CLONAL DIVERSITY OF GROUP A STREPTOCOCCI CAUSING HUMAN INFECTIONS IN SOUTHERN INDIA

A thesis submitted to

The Tamil Nadu Dr. M.G.R Medical University, Chennai

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

By

B. Reena Raj Kumari, M.Sc

DEPARTMENT OF MICROBIOLOGY & VIROLOGY

CHRISTIAN MEDICAL COLLEGE

VELLORE - 632004, INDIA

April 2009

Acknowledgements

I wish to express my deep sense of gratitude to my guide and mentor, Prof. K.N.Brahmadathan, for his guidance, constant support, endless patience and encouragement throughout the doctoral programme. I am grateful to my co-guide Prof. Mary S Mathews for her meticulous approach and invaluable scientific inputs.

Sincere thanks to Dr.Vinodh Abraham, Department of Community Health for facilitating sample collections. Thanks to Mr.Pandiarajan for data management.

Special thanks to Dr.Shaji, Dr. Eunice Sindhuvi, Dr.Priya and Dr.Jeyandran, Department of Hematology for the help rendered during sequencing.

Special word of thanks to my friends and colleagues Ajoy, Aaron, Promila, Solamun, Priya, Ismath, Jonathan, Annie, Angel, Aiswarya, Thelma, Dhinakaran, Charles, Emily and Dhaya for their constant support at work over the years.

I would like to thank my good friend and colleague John Melbin for all his untiring support, encouragement, constructive suggestion and criticism which have helped me to a great extent.

I express my sincere thanks to the entire faculty, microbiologists, other staffs and students of the department of Microbiology, who have helped me in various ways.

I express my deepest gratitude to my parents for their valuable support.

I thank my husband Doss, who has been a pillar of support and encouragement through all this years of study. There are no words to thank my two daughters, Irene and Adrina who waited for me patiently till I returned home…..

Last but not the least, I gratefully acknowledge “The Indian Council of Medical Research” for granting me Senior Research Fellowship.

This work would not have been possible without the blessings of God Almighty who at every stage of this work provided me with strength, perseverance and guidance. To God be the glory Great things he has done…….

CERTIFICATE

This is to certify that the thesis entitled "**CLONAL DIVERSITY OF GROUP A STREPTOCOCCI CAUSING HUMAN INFECTIONS IN SOUTHERN INDIA**" is based on the results of the work carried out by Ms. Reena Raj Kumari for the Doctor of Philosophy degree under our supervision and guidance.

The candidate has independently reviewed the literature, standardized the methodology and carried out the techniques towards completion of the thesis work.

This thesis has not been submitted for the award of any degree or diploma of any other university.

Dr.K.N.Brahmadathan Ph.D
(Supervisor and Guide)
Professor
Department of Microbiology
Christian Medical College, Vellore
Vellore

Dr. Mary S Mathews MD
(Co Guide)
Professor and Head
Department of Microbiology
Christian Medical College,

Station: Vellore

Date

CERTIFICATE

This is to certify that the thesis entitled "**MOLECULAR CHARACTERISATION AND VIRULENCE PROFILES OF GROUP A STREPTOCOCCI CAUSING HUMAN INFECTIONS IN A SOUTH INDIAN COMMUNITY**" is based on the results of the work carried out by Mr. J. John Melbin Jose for the Doctor of Philosophy degree under our supervision and guidance.

The candidate has independently reviewed the literature, standardized the methodology and carried out the techniques towards completion of the thesis work.

This thesis has not been submitted for the award of any degree or diploma of any other university.

Dr.K.N.Brahmadathan Ph.D
(Supervisor and Guide)
Professor
Department of Microbiology
Christian Medical College, Vellore
Vellore

Dr. Mary S Mathews MD
(Co Guide)
Professor and Head
Department of Microbiology
Christian Medical College,

Station: Vellore

Date

Table of Contents

1. INTRODUCTION

GAS, commonly known as Group A Streptococcus (GAS), is a bacterial pathogen that infects the throat and skin and is able to cause a wide range of diseases in humans [1]. This ranges from pharyngitis (strep throat) and impetigo to more serious and fatal diseases such as acute post streptococcal glomerulonephritis, rheumatic heart disease and invasive diseases. Acute rheumatic fever (RF) is an autoimmune sequela of GAS infection which occurs in approximately 0.3- 3% of GAS infected patients aged between 5 and 15 years of all streptococcal sore throats [2]**.** GAS is a major global pathogen with a variety of clinical manifestations. Over 7 million visit pediatricians each year in the United States alone for complaints of "sore throat" and almost a fourth of these visits are due to streptococcal pharyngitis [3]. Acute rheumatic fever and rheumatic heart disease, which are well recognized non-suppurative sequelae of infection by this organism, have devastating public health implications worldwide. An apparent resurgence of acute rheumatic fever in the mid 1980s in the US [4, 5] with ongoing new cases [6] and estimated 10,000 cases of invasive streptococcal disease reported annually in the US [7] confirm that an improved standard of living and use of antibiotic have not adequately reduced the impact of GAS infections. Thus, there is need for a safe, efficacious and cost effective vaccine against GAS infection.

Chronic rheumatic heart disease (RHD) affects mainly the valves of the heart, which become highly deteriorated and surgical valve replacement is warranted [8-11] . The global burden of GAS has a prevalence of 1.78 million new cases of severe GAS disease each year [12**] .** Rheumatic heart disease has a prevalence of 15.6 million cases, with 2, 82,000 new cases and 2, 33,000 death each year [12] . The incidence is more common in developing countries [13] [14, 15]. Shresha and Padmavathi in 1979 found a very high prevalence of

rheumatic fever and rheumatic heart disease, which is 11/1000 among 40,000 school children in Delhi [16] . The long term surveillance study by Grace Koshi (1975-78) among the school children in South India indicated the high prevalence of RF/RHD as 4.9/1000 [17] . GAS infection of the pharynx is usually the precipitating cause of rheumatic fever and pyoderma is usually the precipitating cause for acute glomerulonephritis (AGN) [18]. In India the prevalence of RF/RHD ranges from 0.3 to 5.4 per 1000 children [19] [20, 21]. Oral penicillin V has been used for many years for treatment of RF without change in susceptibility [22-24]. In case of penicillin allergy, erythromycin is the drug of choice. The major challenge for prevention of RF is the development of an effective vaccine [23].

M protein is the major virulence factor of GAS [25]**.**Streptococcal surface proteins exhibit a high degree of antigenic diversity [26, 27]. Analysis of the clonal relationship between clinical isolates from patients by various typing methods is a practical approach to elucidate the epidemiology of the disease [28]. Various methods of phenotyping and genotyping have been described for GAS. Among these methods, M serotyping has been considered the "gold standard" for their characterization [29] [30]. However application of M serotyping is restricted due to lack of a comprehensive set of antisera in most laboratories and the high proportion of non-typeable isolates. The M serotyping has been replaced by sequencing of the 5' coding region of the *emm* gene [31] [32]. This method has been used worldwide as an epidemiological tool to characterize GAS isolates [33]. Currently there are more than 170 *emm* types and 750 *emm* subtypes known. Though *emm* types are good predictors of clonal types within a geographical location, there is considerable genetic diversity within certain *emm* types. Vir-typing and pulse field gel electrophoresis (PFGE) are the other two powerful sub-typing methods for epidemiologic investigation of GAS [30].

DNA-based typing methods are increasingly important for characterization of bacteria, it enables the monitoring of the occurrence, differences and similarities between different groups of bacteria [34]. Data provided from such analyses improves our understanding of bacterial species, genera and populations and provides information for epidemiological and public health interventions [35, 36].

In recent times, epidemiological studies on GAS infection have been carried out using gene sequencing methods. Nucleotide sequence-based methods of GAS allow rapid and global comparison between results from different laboratories. The recent development of multi locus sequence typing (MLST) schemes has provided a new approach to molecular epidemiology that can identify and track the global spread of virulent or antibiotic resistant isolates of bacterial pathogens using the internet [37] [38]. The slow accumulation of variation within housekeeping loci provides framework for long term global studies and reveals information concerning evolutionary relationships [37]. The relatedness among isolates of a bacterial population or species is generally represented as a dendrogram. Similar genotypes generating clusters in the dendrogram displays the evolutionary events very poorly. There are many online softwares for constructing phylogenetic trees but they are inappropriate for exploring the evolutionary relationship among same bacterial species where recombination is frequent [39]. The eBurst algorithm clearly explains and focuses on closely related isolates within a bacterial population [40, 41], which share recent ancestor (clonal complex) [42]. The demonstration of a random association between the alleles at different loci within a bacterial population (linkage equilibrium) strongly suggests a high rate of recombination within the population and departures from linkage equilibrium i.e.

significant levels of linkage disequilibrium [43] [44]between the alleles at different loci are often taken to imply a low rate of recombination.

A thorough documentation of existing GAS clones will lay the ground for gaining a better understanding of the epidemiological trends underlying GAS disease and aid in deciphering the molecular basis for biological diversity within this species [45].Epidemiological data from developing countries for GAS disease is very limited [12] especially from India ,so a thorough documentation of GAS epidemiology based on *emm* typing and multi locus sequence typing will allow one to the monitor the occurrence, differences and similarities between different *emm* types of GAS which will aid in designing vaccines and understand the population genetic structure of GAS.

Hypothesis

Infections due to GAS and their sequelae continue to be an endemic problem in India. Earlier studies from our Laboratory have shown extreme heterogeneity among GAS isolates circulating in our community. This is supported by our preliminary observation that 33 *emm* types were identified among 100 strains isolated from school children during a rural school survey (community strains).

Based on this, we hypothesize that large scale mutation, genetic recombination and natural selection among the vast pool of GAS circulating in the community is responsible for the high heterogeneity. We also hypothesize that allelic variations among certain housekeeping genes of GAS causing infections are responsible for the extreme heterogeneity. To prove this, we plan to use multi locus sequence typing on representative *emm* types identified among rural school children and compare their phylogeny with phylogeny based on *emm* types.

2. AIMS AND OBJECTIVES

2.1. Overall aim

To study the clonal evolution of common *emm* types of GAS causing human infections in a South Indian community using Multi Locus Sequence Typing.

2.2. Specific Objectives

1. To standardize Multi Locus Sequence Typing (MLST) for GAS using seven housekeeping loci for enzymes that performs vital functions.
2. To determine Sequence Types (ST) and their alleles, of GAS causing human infections in South Indian community using MLST.
3. To determine the clonal diversity of common *emm* types of GAS causing pharyngitis, impetigo and clinically invasive disease in South Indian patients by comparing the association of allelic profile of the ST with the diverse *emm* types.
4. To study the linkage disequilibrium within the GAS population by studying the index of association between ST and *emm* types.

3. REVIEW OF LITERATURE

3.1. History of Group A Streptococcus (GAS)

GAS was first described by Billroth in 1874 [46] in patients with wound infections and coined the term ‘streptococcus’ to describe the bacteria. In 1883, Fehleisen [47] isolated chain-forming organisms in pure culture from perierysipelas lesions. Rosenbach named the organism GAS in 1884 [48]. Studies by Schotmuller in 1903 [49] determined that Streptococci produced various kinds of hemolysis on blood agar plates. In 1919 J.H. Brown [50] led to knowledge of different patterns of hemolysis described as alpha, beta, and gamma hemolysis.

A later development in this field was the Lancefield classification of beta-hemolytic streptococci by Rebecca Lancefield and later categorized GAS as specific types based on serotyping of M-protein, this study resulted in the demonstration of serological differences between several of the epidemic strains of GAS [51] . Lancefield established the critical role of M protein in disease causation. At about the same time Griffith differentiated GAS (GAS) into types with slide agglutination test based on T antigens.

In 1920s, Dochez, George, and Dick [52] identified that scarlet fever was caused by hemolytic streptococcal infection. In 1932 Todd [53] developed a test for measuring antibodies to streptolysin O, The epidemiological studies of the mid 1900s helped establish a relationship of GAS infection with acute rheumatic fever (ARF) and acute glomerulonephritis [54].

3.2. Taxonomy and Genus definition

GAS belongs to the family Streptococcaceae [55]. These organisms are gram – positive, catalase – negative, spherical or ovoid cells that tend to grow in pairs and chains, non-motile, non-sporing, fail to reduce nitrate, facultatively anaerobic and requires enriched medium containing blood to grow [56]. The G+C content of DNA is 36-46 % [57]. GAS genome has single circular chromosome in size from 1.85-1.9 Mb [58].The detection of cytochrome enzymes with the catalase test distinguishes members of the family Micrococcaceae (catalase-positive) from the members of the family Streptococcaceae (catalase-negative).

Rebecca Lancefield established the Lancefield grouping system for the beta hemolytic streptococci based on cell wall carbohydrate [51] .These carbohydrate antigens detected on cell wall is composed of repeating units of N-acteyl glucosamine and N-acetylmuramic acid. The Group A polysaccharide is a polymer of N-acteyl glucosamine and rhamnose. This polysaccharide is also known as "C substance or group carbohydrate antigen"[59].

In 1937, Sherman proposed a scheme for placing the streptococci into four categories. These categories were organized by hemolytic reaction, group carbohydrate antigens, and phenotypic tests (primarily fermentation and tolerance test) [60]. Sherman's four divisions were the pyogenic division, the viridians division, the lactic division and the enterococci. The pyogenic division included the β- hemolytic strains with defined group antigens (A, B, C, E, F and G).The nomenclature and taxonomy of streptococci profoundly changed with application of genetic techniques. These include DNA-DNA hybridization, DNA- ribosomal (rRNA) hybridization and small subunit (16S) rRNA hybridization. In 1991, Bentley and coworkers applied partial 16SrRNA sequencing to the streptococcal species to determine the

intrgeneric relationship among them [61]. Kawmura and associates in 1995 extended these studies with some of the newly identified viridians streptococci [62]. Currently GAS is placed in the pyogenic group of streptococcus .The pyogenic group consists of Group A, B, C and G streptococci clustered together phylogenitically, while viridians streptococci fell into several phylogenetic groups.

3.3. Morphology and Cultural characteristics

Streptococci are spherical cells that are arranged in pairs or in chains of up to 50 cells or more. Individual cells are typically 0.5-1.0µm × 1.0-2.0µm.Growth occurs by elongation on the axis parallel to the chain. Cross walls form at right angles to the chain and after division and appearance of pairing may remain. The length varies within wide limits and is influenced by the nature of the culture medium. Chain formation is most pronounced in broth media than solid media. In liquid media, such as glucose, serum broth, Pike's media [63] and Todd-Hewitt broth growth occurs as granular turbidity with a powdery deposit. No pellicles are formed in liquid media.

GAS is a facultative anaerobe growing best at 37°C. It is catalase negative, oxidase negative, non- motile, non-sporing [64]. The minimal nutrition requirements of the streptococci are complex because of the organism's inability to synthesize its required amino acids, purines, pyrimidines and vitamins. GAS is fastidious in their nutritional requirement. For primary isolation of GAS from clinical materials, media enriched with blood is preferred [65]. The optimal pH for growth is 7.4 to 7.6 at 37°C [66]. An enhancement of growth of GAS can be obtained by culture at reduced oxygen tension or an increased level of CO_2. Most GAS is beta hemolytic on sheep blood agar, the presence of small amounts of fermentable carbohydrate (0.05% of glucose) decreases the hemolytic reaction around the surface

colonies. Since hemolysis is enhanced under anaerobic conditions, the agar plates are slashed with the loop at the primary site of inoculation to ensure subsurface growth.

GAS colonies are circular in shape with a diameter range from 0.5 to 2mm. GAS appears in three forms: mucoid, matt and glossy [67]. Mucoid colonies are large and glistening with the appearance of droplets of water. They are viscous in nature and form confluent growth. They are formed by GAS isolates producing large amount of hyaluronic acid capsule. Matt colonies, also known as post-mucoid colonies, are flat with rough surface. These matt colonies are also produced as a result of drying of mucoid colonies. Glossy colonies are smaller than the other forms [68]. They are domed with a shiny surface and are formed by strains that do not produce capsule during growth. The size of the hemolytic zone around surface colonies of GAS is usually 2-4 times the diameter of the colony. In some strains the hemolytic zone is much wider, while in others only a narrow ring is present.

GAS has been reported to occur as nutritional variants (NV) and grow insufficiently or does not grow on the surface of blood agar incubated aerobically [69]. The growth of nutritionally variant Group A Streptococci (NVGAS) can be promoted by incubating in presence of increased CO_2 or addition of special supplementation of the media. The NVGAS grows on blood agar as satellites around normal flora of bacterial colonies [70].

Based on the principle of Pike's medium [71] which incorporates Crystal violet (CV) to inhibit the growth of commensals and increase recovery of GAS from clinical samples, blood agar containing crystal violet has been used for isolation of GAS strains. This is known as crystal violet blood agar. Crystal violet is added in the concentration of 1:1000000(1µg/ml) to 500ml aliquots of blood agar. Crystal violet, a bacteriostatic dye, is used in the preparation of culture media for selective isolation of GAS. This is used in

situation where cultures are overgrown with staphylococci from skin lesions, pharyngeal and nasal cultures. This is a selective agar media for isolation of GAS. Other selective agar media used in isolation of GAS from cultures are Columbia agar with colistin & nalidixic acid and sulfamethaxole- trimethoprim blood agar. Columbia 5% blood agar with colistin (10µg/ml) and nalidixic acid [72] (15µg/ml) inhibits the growth of gram negative bacilli which over grow GAS. A better alternative media for isolation of GAS is Columbia agar with colistin (10µg/ml) and oxolinic acid (5µg/ml) [73]. This inhibits the growth of staphylococci, coryneform and gram negative bacilli. Sulfamethaxole-trimethoprim blood agar is composed of trypticase soy agar with 5% sheep blood supplemented with sulfamethaxole (23.75µg/ml) and trimethoprim (1.25µg/ml). This has been used as the primary blood agar plate which increases the yields of GAS [74, 75].

Specimen collection and processing is the most crucial step in the isolation of ß-hemolytic streptococci (BHS). Various methods, such as transport media (Amies and Stuart), filter paper strips [76] and silica gel desiccated swabs [77], have been used to improve the survival of BHS [78], especially GAS, and prevent growth of commensals. Filter paper strips measuring 2 x 6 cm impregnated with crystal violet used in the transport helps in the recovery of GAS from clinical samples. This is a good technique for isolation of GAS the epidemiological study of streptococcal pyoderma [79]. GAS isolates can be preserved by lyophilisation, sand desiccation method and Robertson cooked meat medium [80].

3.4. Structure, Antigenic composition and Virulence factors of GAS

The cell surface of GAS accounts for many virulence factors those concerned with colonization and evasion of phagocytosis and the host immune response [81] . The outer most layer is the capsule consisting of hyaluronic acid [82] . The cell wall surface is covered

with hair like protrusions (made up of M, T and R protein antigen), fibronectin binding proteins, cell bound streptokinase and lipoteichoic acid [83]. The cell wall consists of peptidoglycan made up of repeated units of N-acetylglucosamine and N-acetylmuramic acid cross linked by a short peptide [84, 85]. GAS produces a number of toxin and enzymes. Among these are superantigen streptococcal pyrogenic exotoxins (erythrogenic toxins A, B, C), Streptolysin S & O and other extracellular antigens such as deoxyribonuclease B (DNase B), hyaluronidase, streptokinase and nicotinamide adenine dinucleotidase (NADase) [86].

The mechanisms by which GAS initiates invasive infections are currently a focus for numerous research groups world-wide. Several virulence determinants which aid the pathogen in evading the immune system, adhering to the host epithelial cells and invading the host tissue have been identified. The advent of molecular techniques has revealed much about the organism's virulence and the genomes of several GAS types have been deciphered [87]. Surface structures of GAS including a family of M-proteins, the hyaluronic acid capsule and fibronectin–binding proteins, allow the organism to adhere, colonise and invade tissues [88, 89]. M protein binds to complement control factors and other host proteins to prevent activation of alternate complement pathway and thus evade phagocytosis and killing by polymorphonuclear leucocytes [90]. Extracellular toxins, including superantigenic streptococcal pyrogenic exotoxin contribute to tissue invasion [1].

3.4.1. Hyaluronic acid capsule

The GAS capsule is composed of hyaluronic acid, a high molecular- weight polymer [91] . The hyaluronic acid capsule is required for resistance to phagocytosis [92, 93]. The production of hyaluronic acid is controlled by an operon composed of three different genes *hasA, hasB* and *hasC* which are transcriptionally regulated. *hasA* encodes hyaluronate

synthase , *hasB* encodes for UDP – glucose dehydrogenase and *hasC* encodes UDP-glucose pyrophosphorylase [94]. The inactivation of *hasC* does not affect the production of capsule [95]. GAS strains rich in M protein and capsules are extremely mucoid and are highly virulent. GAS capsular hyaluronate is very similar chemically to the connective tissue of human [96]. GAS hyaluronic acid capsule is a poor immunogen, so there is insignificant antibody production in humans [97, 98].

These mucoid strains with capsules have been found associated with virulence has been supported by the observation that the capsule is formed by 3% of isolates from patients with uncomplicated pharyngitis, 21% of isolates from severe streptococcal infections and 42% from acute rheumatic fever which indicates the importance of capsule in invasiveness [99]. Among the frequently isolated serotype M1 the prevalence of capsule production is even higher (6, 22 and 80% respectively). Capsule producing strains are highly virulent in animal models.

3.4.2. Cell wall

The cell wall consists of peptidoglycan back bone with integral lipoteichoic components. The main function of these components is structural stability of GAS [100-102] . Lipoteichoic acid (LTA) facilitates the adherence of GAS to pharyngeal epithelial cells [103, 104]. Peptidoglycan, like endotoxin of gram negative bacteria, is capable of activating the alternative complement pathway [105-107]. Lipoteichoic acid serves by hydrophobic interactions as a first step adhesin, bringing the organisms into close contact with host cells and allowing other adhesions to promote high affinity binding [89].

GAS surface proteins that bind fibronectin have been studied extensively and are important in adherence to both throat and skin [108-110]. These include protein F1 also known as SfbI (streptococcal fibronectin binding protein I) [111]. Other proteins which are found in relation to SfbI are SfbII, FBP54 (fibronectin-binding protein), protein F2 and PFBP (*Streptococcus pyogenes* fibronectin-binding protein**)** [112]. Protein SfbI facilitates adherence to pharynogeal epithelial cells[113]. Mouse challenged with lethal dose of virulent GAS was found to be protected when vaccinated with intranasal vaccination of protein SfbI [114-116].

3.4.3. M protein

The ability of GAS to persist in infected tissues is due primarily to the cell surface protein exposed M protein [90] , a molecule which confers GAS ability to resist phagocytosis by polymorphonuclear leucocytes in the absence of specific antibody. Antibodies to M protein are associated with type specific immunity. Resistance to infections by GAS appears to be related to presence of antibodies in secretion and sera to the M protein molecule [117]. M protein is one of the main virulence factors. Under the Lancefield classification scheme there are, nearly 80 antigenic M protein serotypes.

Structural studies of the streptococcal M protein began in the late 1970s and early 1980s by Beachey and colleagues [118] and Manjula, Fischetti, and colleagues [119] when the protein chemistry was used to obtain the amino acid sequence of peptide fragments of the M protein. The structural basis of the biological properties of M protein was elucidated using M type 24, 5 and 6 [118, 120, 121]. M protein is composed of two polypeptide chains complexed in an alpha helical coil configuration anchored in the cell membrane, traversing the cell wall and appearing as fibrils on the cell surface [122] . The C terminus of the coiled M protein acts as a membrane anchor. From the membrane anchor region towards the amino

(NH_2) terminus, the M protein is composed of proline/ glycine – rich region embedded in the cell wall followed by repeated regions of repeating amino acids designated A-D. The amino acid sequences of regions B-D are highly conserved among different M proteins. The A repeat region and the N (amino) terminal portion of the molecule extends into the environment and concludes in a series of approximately 11 amino acids. This short sequence and the region A constitute the hypervariable region of M protein. Antigenic variation in the hypervariable region constitutes the basis for the Lancefield serological classification of GAS [123].

Cloning of the M protein gene has expanded the M typing. The M protein gene (*emm* gene) added more *emm* types to the collection which was not possible by sero-typing. Currently there are more than 125 M protein types [124] and 175 *emm* types with 750 *emm* subtypes [125].

Cross-protection between M serotypes was first reported in 1939 by Hirst and Lancefield, who immunized mice with crude acid extracts of M protein [126, 127]. Cross reaction among M types were correlated with the presence of common peptide fragment which suggested a certain degree of structural similarity among cross reactive serotypes of M protein. Cross reaction to such shared determinants may also account for the limited degree of heterologous protection reported by Wittner and Fox [128] who immunized mice with a trivalent vaccine that contained acid extracts of type 1, 3 and 12 M proteins. In addition to protection against vaccine serotypes, the immunized animals were also protected against challenge infections with M type 6 and 14.

M proteins have been divided into two classes, class I and class II molecules based on their reaction with antibodies against the C repeat region of M protein. Class I M protein is reported to contain a surface exposed epitope on the whole GAS that reacts with the antibodies against the C repeat region [129]. GAS containing class II M protein does not react with antibodies and do not contain class I epitope. The class I M protein serotypes have serum opacity factor while class II M protein do not [130]. In some studies, there was strong correlation between serotypes known to produce rheumatic fever and the presence of class I epitope [131].

Immunity to M protein has shown to be protective against GAS infections and paved a way for vaccine studies based on M protein [132]. The immune response to M protein is protective (by production of protective antibodies which promote phagocytosis) and destructive (antibodies react to the host tissues) [133].

Genes related to M protein gene are called the M (*emm*) gene superfamily [134]. They include immunoglobulin binding proteins, M- related proteins and M proteins. According to Hollingshead and colleagues more than 20 genes have been identified in the *emm* gene superfamily [134] in which genes shared more than 70% DNA sequence identity at their 5'end [135]. The genes in the *emm* family are located in a cluster occupying 3-6 kb between the regulatory genes *mga* and *scpA* gene encoding C5a peptidase within the *vir* regulon on the chromosome of GAS [136]. Five major *emm* chromosomal patterns of the *emm* subfamily genes were identified based on the number and arrangement of *emm* subfamily genesc[137] **.** These subfamily arrangement were designated A to E.

The immunoglobulin binding proteins identified for GAS are encoded by *emm* or *emm* related genes. They are immunoglobulin binding proteins which have structural similarity to M related proteins and interact with immunoglobulins outside their antigen combining site in the F_c region of the immunoglobulin molecule [138]. Heath and Cleary cloned and sequenced the first immunoglobulin binding protein gene, the *fcrA* gene. Genetic studies have suggested that a common ancestral gene has undergone gene duplication to produce the diversified family of immunoglobulin binding proteins [135, 139-144]. The genes encoding the immunoglobulin binding proteins are controlled by multiple gene regulators (*mga*) of GAS, a positive transcriptional regulator also controlling expressions of M protein.

Antigenic heterogeneity of M protein is the consequence of highly polymorphic 5'ends of *emm* family genes that encodes amino acids distal to the cell surface capable of provoking a protective immune response during infection**.** The 5'end is composed of repeats that participate in intragenic recombination resulting in size variation of the M protein.

3.4.4. Serum opacity factor

Top and Wannamaker demonstrated that production of opacity factor is closely related to M types [145]. Approximately half of GAS strains are capable of opacifying mammalian serum, a unique phenotype attributable to the presence of serum opacity factor (SOF) encoded by the *sof* gene [146, 147]. SOF is a 110-kDa protein with an LP(X)SG cell wall anchor motif expressed in both membrane bound and extracellular forms [148-150]. SOF is a bifunctional protein comprised of an N-terminal opacification domain, and C-terminal domain of highly conserved tandem repeat sequences capable of binding the host extracellular matrix components fibronectin and fibrinogen[150-152] .Once thought to be a lipoproteinase, SOF has recently been shown to bind apoA-I and apoA-II, causing release of

HDL lipid cargo, which in turn coalesces to form lipid droplets, resulting in serum opacification [153].Studies with SOF-coated latex beads indicate the protein can promote adherence to human HEp-2 pharyngeal epithelial cells including fibronectin-specific interactions [154]. A role in GAS virulence has been attributed to SOF, since insertional inactivation of the *sof* gene in a serotype 2 GAS isolate reduced mortality in a mouse intraperitoneal infection model [148]. SOF also triggers host production of opsonic antibodies that protect against infection by SOF (+) strains of GAS, suggesting it may represent a useful vaccine antigen [130] .

The *sof* gene is recently recognized to be a member of a two gene operon, invariably cotranscribed with a gene encoding a second surface anchored protein, *sfb*X [155]. SfbX possesses a C-terminal domain of tandem repeat sequences very highly homologous to the fibronectin-binding domain of SOF, but a unique N-terminal domain. SfbX is capable of binding fibronectin, but the function of its N-terminal sequences has yet to be elucidated [155]. Given previous insertional mutagenesis studies targeting *sof* were performed before the operon structure was recognized, polar effects on *sfb*X would be anticipated, and the precise contribution of each individual gene is therefore uncertain.

3.4.5. C5a peptidase (SCPA)

C5a peptidase is a proteolytic enzyme (endopeptidase) found on the surface of GAS [156] . It is highly specific 130kDa serine peptidase that is anchored to the streptococcal cell wall [157].The SCPA gene is encoded by a gene which is regulated by *mga* in concert with M protein [158]. Its role as major virulence factor has been established**.** The enzyme SCPA is produced by group A, B and G streptococci [1] . SCPA cleaves the complement–derived chemotaxin, C5a at the His^{67}-Lys^{68} bond, thus removing the leucocyte-binding site of C5a

[156]. As a result SCPA inhibits chemotaxis, delaying infiltration of phagocytosis, thus promoting the infection of GAS by establishing at the site of infection [159, 160]. Animal studies using SCPA as an immunogen has been done to check for protection rendered by SCPA against GAS infection. Intranasal immunization of mice with recombinant SCPA induced strong humoral and mucosal response and reduced nasopharyngeal colonization after challenged with M types suggesting cross protection [161]. Antibody against SCPA was found in detectable levels against uninfected adults rather than children who are healthy [162] .

3.4.6. Streptokinase

Streptokinase is a 46 kDa single chain, 414 amino acid protein, secreted by group A, C and G streptococci[163] . Crystal structure studies of the activator complex of streptokinase with microplasmin (the catalytic domain of plasmin) indicate that streptokinase contains three domains (α, β and γ) which bind to the catalytic domain of plasmin. The streptokinase gene is highly conserved [164].

Streptokinase interacts with plasminogen to form streptokinase - plasminogen complex with proteolytic activity capable of converting other plasminogen molecules to plasmin. The breakdown of tissue barriers by plasmin facilitate bacterial spread and also supply oligopeptides and amino acids to be used by the pathogen [165] . Attention towards streptokinase has been offered due to clinical use as a thrombolytic agent [166] . More interest in streptokinase has been seen because of its association with acute post-streptococcal glomerulonephritis [99] .

The complete nucleotide sequence of the streptokinase gene has been reported earlier for GAS M type 49[167] and M type1[168]. Sequence comparison analysis of streptokinase

proteins from different strains of streptococci revealed marked differences in the degree of similarity .Streptokinase from GAS M type 1 and group C and G formed one cluster and GAS M type 12 and 49 made up another variant [169, 170] . The nucleotide identity between nephritis and non-nephritis GAS isolate is 90% and 85 % at amino acid level**.** The non-identical amino acids are located in two major variable regions V1 and V2 corresponding to amino acid residues 174-244 and 270-290 respectively .The V1 regions of nephritis associated with *ska* allele show more than 95% homology with each other at nucleotide level but less than 60% homology with the non-nephritis *ska* allele [171, 172] .

Synthesis of the plasminogen activator streptokinase (SK) by GAS has shown to be subject to control by two two-component regulators, *covRS* (or *csrRS*) and *fasBCA* [173]. In independent studies, response regulator CovR proved to act as the repressor, whereas FasA was found to act indirectly as the activator by controlling the expression of a stimulatory RNA, *fasX*.

3.4.7. Streptolysin O and S

Streptolysin O is a protein of approximately 60 kDa. It belongs to a family of oxygen labile; thiol activated cytolysins and causes the broad zone of beta hemolysis surrounding colonies of GAS with cut on blood agar plates [174]. Thiol activated cytolysins bind to cholesterol on cell membranes, creating toxin-cholesterol aggregates that contribute to cell lysis via a colloid –osmotic mechanism. Cholesterol inhibits both toxicity in isolated myocytes and hemolysis of red blood cells *in vitro*. In conditions, where serum cholesterol is high (nephritic syndrome), falsely elevated ASO titers may occur because both cholesterol and anti-ASO antibody will neutralize streptolysin O. They are antigenically related, although the similarities are not evident from sequence comparison of the respective genes encoding

these toxins. Striking amino acid homology exists between streptolysin O and thiol activated cytolysins from other gram- positive bacteria. Intra venous injection of streptolysin O into mice, rabbits, guinea pig causes death within seconds [175]. Death is the result of an acute toxic action on the heart. Serum antibodies to streptolysin O can be demonstrated, the highest titers are found in patients with rheumatic fever [176]. Streptolysin S is a cell associated hemolysin that does not diffuse into the agar media. It is produced by GAS in the presence of serum albumin, alpha-lipoprotein, and ribonucleic acids. Its hemolytic activity is inhibited by serum lipoprotein and other phospholipids. It is not antigenic. Unlike SLO it is not inactivated by oxygen. It is thermo labile. Synthetic peptides containing aminoacid residues of the SLS molecule evoke toxin neutralizing antibodies [177].

3.4.8. Streptococcal pyrogenic exotoxin (Spe)

Pyrogenic toxins are a group of acid-stable single-chain proteins ranging in molecular weight from 20,000 to 40,000. This group includes the staphylococcal enterotoxin serotypes A to E (SEA to SEE), the staphylococcal pyrogenic exotoxins A and B, toxic shock syndrome toxin type 1(TSST-1) [178], and the streptococcal pyrogenic exotoxin types A to C (SPE A to SPE C). These toxins are characterized by their similar biological activities: they induce fever, are nonspecific T-cell mitogens, enhance host susceptibility to endotoxin, suppress B-lymphocyte function, and cause a scarlet fever-like rash. Some members of the group, such as SPE A, staphylococcal enterotoxin SEB, and staphylococcal enterotoxin SEC, share significant sequence similarity [179].

The SPEs, formerly called the scarlet fever toxins, have been shown to be important in the development of scarlet fever. The speA and speC genes are encoded by bacteriophage [180-183]while the SpeB is chromosomal [184-186]. SpeB is a cysteine protease which degrades

vitronectin, fibronectin and IL-1 precursor [187, 188]. Other pyrogenic exotoxins include mitogenic factor (SpeF), streotococcal super antigen (SSA), streptococcal mitogenic exotoxin Z (SMEZ), *speG, speH, speJ* and *speK* [189, 190].

spe genes – *speA, speB, speC, speF, speG, speH, speJ, speK,ssa,* and *smeZ* encode a family of high mitogenic proteins secreted by GAS [191]. Most of the genes exhibit high superantigenic activity. Among the *spe* genes, *speB* is exceptional because its mitogenic activity is less compared to the other genes because it is a cysteine protease. SMEZ was distict from other known mitogenic exotoxins and anti –SMEZ antisera did not cross react with SpeA, SpeB and SpeC in neutralization tests of mitogenic activity.

The nucleotide sequence of *speA* gene was found expressed among GAS isolated from streptococcal toxic shock syndrome and it accounted to 85% of the isolates. Nucleotide sequencing of streptococcal pyrogenic exotoxin A gene from outbrakes revealed four alleles of *speA* (*speA1, speA2, speA3* and *speA4*). *speA1, speA2* and *speA3* differ from each other by single amino acid while, *speA4* encodes toxin which differs by 9% divergent from the three and *speA4* has 26 amino acid substitutions [192]. The allele *speA3* is highly mitogenic and has more affinity for the HLA DQ molecule. Inflammatory response triggered by the streptococcal pyrogenic exotoxins or superantigen lead to autoimmune responses and manifestations[193].

3.4.9. Cysteine protease (SpeB)

Streptococcal proteinase (SpeB) is an extracellular cysteine proteinase which is a 40 kDa zymogen which is cleaved into 28kDa mature enzyme under reducing conditions [185, 194, 195]. Increased antibody production is seen in cases of pharyngitis, rheumatic fever and

invasive cases against SpeB, while low levels of antibodies are seen in fatal invasive disease of GAS [196].

SpeB has been shown interact with many host proteins. It cleaves several proteins such as fibrenectin and vitronectin and facilitating bacterial dissemination, colonization and invasion [187]. Several *in vitro* studies to analyse the biological importance of SpeB secretion in GAS disease has been carried out. SpeB has been detected *in vitro* during the stationary phase of growth in GAS and its production has been inhibited in a nutrient rich environment [197].Thus SpeB secretion has been illustrated as an adaptive mechanism of the pathogen in a nutrient deficient environment. SpeB is also known to degrade streptokinase in vitro in late stationary phase of GAS growth. The addition of cysteine protease inhibitor prevented the degradation by M 53 type strain [198]. Proteomic analysis of the culture supernatants of an M1 type GAS isolated from invasive GAS disease has shown that SpeB is capable of degrading most of the GAS secreted proteins including M protein and streptokinase [199]. This degradation has also been associated with decrease in mitogenic activity of lymphocytes *in vitro*. The *in vitro* studies of M1 GAS strain cultured in human saliva have shown high density colony forming units in the stationary phase. The ability of strain to maintain such a growth pattern has been attributed to number of virulence factors including SpeB. This study helps in gaining an insight into the molecular process that enhances colonization in the upper respiratory tract.

Transcription of *speB* has shown to be controlled by the gene *oppA*, which is a part of polycistronic operon. The isogenic mutant of *oppA* has been shown to decrease transcription of *speB* and cause less mortality and tissue damage in mouse model [200]

3.5. Clinical manifestations

GAS is the etiologic agent of streptococcal pharyngitis and streptococcal impetigo, which result in serious disease sequelae such as acute rheumatic fever (ARF), rheumatic heart disease (RHD) and acute post streptococcal glomerulonephritis (APSGN). It is also responsible for invasive disease such as streptococcal toxic shock syndrome (STSS), necrotizing fasciitis (NF), myositis and bacteremia [1, 201].

3.5.1. Non-invasive streptococcal disease

3.5.1.1. Pharyngitis

The most common infection due to GAS is streptococcal sore throat. GAS adheres to the pharyngeal epithelium by means of lipoteichoic acid covered surface pili. The glycoprotein fibronectin on epithelial cells serves as lipoteichoic acid ligand. It primarily affects the school children aged 5-15 years [202]. Clinical signs such as abrupt onset of sore throat, malaise, and headache generally develop 2-4 days after exposure. The posterior pharynx is usually reddened, with enlarged tonsils. The local inflammation results in swelling of cervical lymph nodes. Occasionally tonsilar abcess develop which may spread to neighbouring regions and to the bloodstream.

Pharyngitis caused by certain pyrogenic exotoxin of GAS may be associated with diffuse erythematous rash of the skin and mucous membrane. This condition is known as scarlet fever [203]. The rash develops within 1-2 days after first symptoms of pharyngitis and initially appears on the chest, then spreading to the extremities. The tongue becomes red and denuded (strawberry tongue).

3.5.1.2. Impetigo

GAS causes several skin infections, sometimes in association with *Staphylococcus aureus.* The superficial and localized skin infection known as impetigo or pyoderma occurs mainly in children. It primarily affects exposed area on face, arms or legs. The skin becomes colonized after contact with infected person and the bacteria enter through small defects [204]. Initially, clear vesicles develop, which within few days become pus filled. Secondary spread is often as a result of scratching.

More severe acute infection is the acute skin infection erysipelas, which occurs in the superficial layers of the skin (cellulitis) and involves lymphatics[204]. The infection is characterized by diffuse redness of the skin, and patient's experience local pain, enlargement of regional lymph nodes and fever. Untreated, the infection may spread to the blood stream.

3.5.2. Invasive soft tissue infection

GAS invasive infections are severe and life threatening [205]. The most severe forms of invasive infections are necrotizing fasciitis (NF), streptococcal toxic shock syndrome and puerperal fever [206], all of which are associated with bacteraemia. Once in the blood GAS multiplies with incredible speed and the mortality rate reaches 40%.The potential complication include acute endocarditis leading to heart failure.

3.5.2.1. Necrotizing fasciitis (NF)

Necrotizing fasciitis (NF) disease is a deep seated infection of the subcutaneous tissue that results in the progressive destruction of fascia and fat. This was previously known as streptococcal gangrene [207, 208].GAS gain entry to these tissues through trauma .Systemic shock and general deterioration occur very rapidly. The clinical diagnosis may be difficult

because *Staphylococcus aureus* and anaerobes such as *Clostridium perfrigens* can produce similar clinical picture. GAS can be identified from the blood, blister fluid and cultures of infected area.

Predisposing factors for the development of NF due to GAS include varicella, penetrating injuries, minor cuts, burns, splinters, surgical procedures, child birth, blunt trauma and muscle strain [209]. The first cutaneous clue to streptococcal necrotizing fasciitis is diffuse swelling of arm or leg followed by the appearance of bullae filled with clear fluid which rapidly takes on a maroon or violaceous colour. Unless appropriate intervention is undertaken, there are chances for cutaneous gangrene, with myonecrosis and an extension of the inflammatory process along fascial planes. There are marked systemic symptoms, which may include shock and organ failure [209].

3.5.2.2. Streptococcal toxic shock syndrome

The early onset of shock and isolation of GAS from a normally sterile site are the defining characters of Streptococcal toxic shock syndrome[210]. Patients with invasive and bacteraemic GAS infections and in particular necrotizing fasciitis may develop Streptococcal toxic shock syndrome. The disease, which was first described in the late 1980s, is the result of streptococcal toxins to the bloodstream. The early symptoms of streptococcal shock syndrome include myalgia, malaise, chills, fever, vomiting and diarrhea. Without treatment, the disease progress to shock and organ failure.

3.5.3. Non -suppurative sequelae

Two serious diseases develop as sequelae of GAS infection. They are rheumatic fever (RF) and acute post streptococcal glomerulonephritis (APSGN). Both are caused by immune

reaction induced by streptococcal infection. The first clinical sign appears 1-5 week after the infection and at the time the bacteria have been eradicated by the immune system or as a result of antibiotic therapy.

3.5.3.1. Rheumatic fever (RF)

Rheumatic fever manifest as an inflammation of joints (arthritis), heart (carditis), central nervous system (chorea), skin (erythema marginatum) and subcutaneous nodules. Acute RF is an inflammatory disease of childhood that occurs following untreated GAS pharyngitis infection. Rheumatic fever is diagnosed by physician based on the presence of criteria found in the revised Jones criteria [211].

Rheumatic fever is a major cause of acquired heart disease in young people throughout the world. The incidence of Rheumatic heart disease worldwide ranges from 0.5 to 11 per 1000 of the population [212]. Rheumatic fever is autoimmune in nature and results from the production of auto reactive antibodies and T lymphocytes induced by cross reactive components of the bacteria and host tissues (molecular mimicry) [1]. The medical importance of rheumatic fever is serious cardiac involvement, with myocarditis or valvulitis leading to death or valve replacement.

The strongest evidence supporting the hypothesis that rheumatic fever is a result of GAS infection is that rheumatic fever (RF) parallels the occurrence of streptococcal infection. In addition, elevated antistreptococcal antibodies accompany rheumatic fever, such as rise in ASO and anti-DNASe B antibody titers. Rheumatic fever has marked tendency to be reactivated by recurrent streptococcal infections, whereas APSGN does not.

3.5.3.2. Acute post streptococcal glomerulonephritis (APSGN)

Acute glomerulonephritis is most often seen in children may be associated with acute onset of edema, oliguria, hypertension and hematuria. Various mechanisms involved in the development of APSGN are immune complex deposition of GAS, reaction of antibodies cross reactive with streptococcal glomerular antigens, direct toxic effect on glomerulus, and alteration of glomerular tissues by streptococcal toxins such as proteinase or streptokinase, direct complement activation by streptococcal complements deposited in the glomeruli. Kefalides and colleagues demonstrated that sera from patients with APSGN contained antibodies against lamnin, collagen and other macromolecules found in the glomerular basement membrane. It was also found that the epitope recognized in collagen resides in the 70s domain of the type IV collagen [213].

The typical clinical feature of APSGN includes edema and smoky dusty urine. Patients also exhibit pallor and may contain lethargy, malaise, weakness, anorexia, headache, and dull back pain. Fever is not prominent. Facial and periorbital edema are usually present but also involves foot, leg, scrotum and sacrum. Another manifestation of fluid overload is circulatory congestion, which may give rise to dyspnea, orthopnea, rales at the lung bases, distended neck veins and even pulmonary edema.

3.6. Laboratory diagnosis

In acute infections, diagnosis is established by culture, while in the non suppurative complication the diagnosis is mainly based on the demonstration of antibodies. A microbiological diagnosis is essential for throat culture because all the sore throats are not caused by GAS, it could be viral. If the sore throat is caused by GAS it is essential to

administer penicillin in prevention of primary attacks and recurrences of rheumatic fever. This has been demonstrated by Wannamaker *et.al.*, Siege *et.al.* and Wood *et.al.*

3.6.1. Collection of specimen

The diagnosis of GAS infection is established by demonstrating the pathogen in throat or skin swabs, pus, blood cultures, CSF, expectorates or urine according to the site of infection.

3.6.2. Cultivation and identification

The specimen collected in aseptic precaution is plated on agar containing 7% sheep blood and incubated at 37°C anaerobically or 5-10% CO_2. The hemolysis develops better under this condition. Slicing the innoculum into the blood agar has similar effects. The colonies are about 1mm in diameter and form large zones of hemolysis.

Presumptive information may be obtained on Gram stain from specimens. The presence of gram positive cocci in chains is indicative of GAS infection. However smears are of no value in infections of throat or genitalia, where streptococci may form resident flora.

3.6.3. Bacitracin susceptibility

Bacitracin susceptibility test is the method of choice to identify GAS. This is a disc diffusion method using bacitracin discs with concentration 0.04 to 0.05 international units. Zone of inhibition greater than 14 mm in diameter are given by GAS. This test has a sensitivity of >95%, is only a presumptive test and is not recommended since group G and C streptococci can give false positive results [214]. Batch to batch variation may occur in the commercial discs and therefore it is essential to test each batch for quality control with known GAS strains. If this is done regularly, one can resort to GAS identification by this method.

3.6.4. Group carbohydrate identification

The recommended method of GAS identification is by testing β -hemolytic colonies on BA for group A specific carbohydrate antigen (ACHO) [215]. Numerous methods are available in the laboratory for this, of which the time tested Lancefield's hot-acid extraction technique and Fuller's formamide extraction method are the most widely used.

The micro-nitrous acid extraction procedure of El Kholy is a rapid method. When performed together with the co agglutination reagents, this technique can identify a GAS strain from a BA plate in about 30-45 minutes [216].

Twenty microliter of 2M sodium nitrite solution was taken in a test tube. Three to four colonies of β -hemolytic streptococci (BHS) was inoculated into this. Three μl of glacial acetic acid was added to the suspension and left at room temperature for 15 minutes. Sixteen to twenty four μg of sodium bicarbonate was added for neutralization, after which 60 μl of distilled water was added. This is the antigen extract.

The reagent is composed of suspension of *Staphylococcus aureus* (Cowan I strain) which has cell wall associated protein A that can bind to F_c portion of IgG molecules leaving the F_{ab} portion free for binding. Using a micro pipette, one drop of well mixed antigen extract was added into ring of a ceramic slide. One drop of Group A co agglutination reagent (Staphylococcal suspension sensitized with group A antiserum) was added to this.

The slide was rotated for 3 minutes to mix the suspension well. Agglutination was observed against diffuse light. Development of clumping and clearing confirmed the test strain as GAS. If agglutination was weak (1+), the co agglutination was repeated with a 6 hour old culture of the test strain in THB.

3.6.5. Rapid antigen detection test (RADTs)

Rapid antigen detection test have a basis of extracting carbohydrate antigen from BHS and identify them by immunological methods. Despite their higher cost, they can provide a result within a few hours even when done directly on throat cultures [22] .This enhances the patient compliance to treatment and has been shown to significantly increase the number of patients appropriately treated for GAS pharyngitis [217]. Although tests such as latex agglutination, co-agglutination, enzyme immunoassays, liposomal and optical immunoassays have been employed for this purpose, most of them have sensitivity between 70% and 90%, but with a specificity of >95% with culture as a gold standard [22]. Therefore a negative RADT should always be confirmed with a throat culture. Though molecular techniques have been standardized recently with improved sensitivity, factors such as cost and feasibility has hindered their use in smaller laboratories.

3.6.6. Serological Diagnosis

Determination of anti-streptolysin O antibodies (ASO) had been the mainstay of confirming a diagnosis of GAS pharyngitis as well as RF [218]. Demonstration of a significant or four-fold rise in titer on paired serum samples taken at an interval of 7 to 14 days apart will indicate an ongoing or an acute infection. On the other hand, presence of GAS in throat in the absence of a significant rise in antibodies indicates a carrier state and no GAS infection. Practical difficulties in getting two serum samples from children and the time taken to demonstrate a four-fold rise in titer make this unfeasible on a routine basis. Alternately, titer obtained with a single serum sample can be interpreted based on a cut-off value defined as the upper limit of normal (ULN). ULN represents the highest level of antibodies that can be observed in 20% of normal individuals who have demonstrable antibodies in them. It is to be emphasized that ULN titers should be determined in different geographical areas because

such titers are distinctly higher in endemic regions as compared to non-endemic areas. Any ASO titer above these cut-off values will indicate a GAS infection. This method is generally convenient and reliable although an antibody response from an earlier GAS infection may confuse the final interpretation of the current ASO titer.

One major disadvantage of ASO test is its inability to demonstrate an antibody response in many cases of impetigo. In such cases, determination of anti-deoxyribonuclease B antibody (ADNB) is a valuable method of diagnosing GAS infection [219]. The ADNB test is more sensitive and the magnitude of antibody response is higher than that of ASO both in cases pharyngitis and impetigo as well as the non suppurative sequelae.

Determinination of ASO and ADNB titers by the conventional micro-titer technique using SLO and DNase B enzyme produced and standardized in our laboratory [220]. These tests are labor intensive, cumbersome and time consuming. Therefore, nephelometric titration (Nephlometer), which is an automated, simple and rapid method which does not require much technical expertise, is in use. Many commercial latex agglutination-based ASO kits are available for diagnostic purpose [21].

3.6.7. Antibiotic Susceptibility Testing

Antibiotic susceptibility may be performed by either dilution or diffusion methods. The choice of method is often based on the ease of performance, availability of materials and flexibility of method. While disc diffusion is simpler method, it provides only qualitative results.

Penicillin remains the drug of choice for the treatment of GAS infections and their sequelae with macrolides recommended for patients allergic to beta-lactams. However, since the first

description of erythromycin-resistant GAS in 1955, many countries have reported varying frequencies of erythromycin resistance, with a significant global increase in the incidence of resistance in recent years .This increase in resistance is mediated either through target modification caused by methylase activity encoded by two classes of methylase genes, *ermAM* and *ermTR*, or by an active drug-efflux pump encoded by the *mefA* gene .The increased use of macrolides for the treatment of upper respiratory tract infections has contributed greatly towards the development of resistance [221]. Continuous surveillance of the susceptibility patterns of GAS should therefore be performed in order to recommend appropriate guidelines for antibiotic prescription.

Despite reported penicillin tolerance, GAS continues to be exquisitely susceptible to penicillin which therefore is the drug of choice to treat the infections caused by them. Thus in practice it is not necessary to test their susceptibility on a routine basis; however it is imperative to look for emergence of resistance to penicillin, locally. More importantly, resistance to other antibiotics such as macrolides [222] is on the increase which can be monitored only by regular testing of GAS strains to such antibiotics. Antibiotic susceptibility is performed on Muller Hinton blood agar for GAS.

3.7. Epidemiology

Severe invasive infections caused by GAS have been reported with increased frequency in United States and from other parts of the world. This increase in the incidence of GAS infections has frequently been associated with specific clones, suggesting the possibility that the rise of particular virulent clones is responsible for the reemergence. The identification of GAS clones in the surveillance and epidemiological studies has frequently relied on serotyping using two variable surface antigens, the T antigen (T typing) and the M protein

(M serotyping). Serology based typing methods have their own limitations. Typing methods are the major tools for the characterization of bacterial pathogen allowing the determination of the clonal relationship between isolates. Recent technological advances have resulted in a shift from classical phenotyping typing methods such as serotyping, biotyping and antibiotic resistant typing to molecular methods such as restriction fragment length polymorphism, pulse field gel electrophoresis, PCR based methods, PCR serotyping, sequence based typing methods and microarray.

Clonally related isolates share characteristics by which they can be differentiated from unrelated strains. In epidemiological terms, clonally related organisms are members of the same species that share virulence factors, biological traits and genomic characteristics. This ability to discriminate between genome is essential to several disciplines of microbiology including taxonomy, studies of evolutionary mechanisms, phylogenetic relationships and population genetics of organisms. There are many typing techniques which provide good molecular differentiation and which can be tailored to suit the needs or the nature of the clinical study. However, each of these methods has technical and target dependent limitations that should be considered before the choice of a typing technique is made. Several criteria are proposed for evaluating the performance of typing system, they are typeability, reproducibility, stability, discriminatory power and epidemiological concordance. The following criteria can be considered for convenience; flexibility, rapidity, accessibility and ease of use based on factors like the scale of epidemiological investigation, the time when the information is required, the financial aid and technical resources.

3.7.1. Typing Methods

3.7.1.2. *emm* typing

The first classification for streptococci was based on the hemolytic activity of isolates. Subsequently, the Lancefield classification scheme of serologic typing distinguished the streptococci based on the group A carbohydrate present in the cell wall. GAS was further classified into different M serotypes based on M surface protein. More than 80 different M protein serotypes have been identified using this method where the surface protein can be extracted from the bacteria with boiling hydrochloric acid. Serotyping faced several problems, to avoid these problems; a molecular approach for the identification of *emm* genes encoding the M protein in GAS was used. Currently there are more than 170 *emm* types and 750 *emm* subtypes known from GAS.

Approximately 40% of GAS isolates are capable of opacifying sera due to the expression of the serum opacity factor gene (*sof*).Reported literature of Beall based on *emm* sequence and *sof* gene sequence variation highlighted that within an *emm* type with same specific M type had differing genetic background based on *sof* gene and T agglutination patterns. Beall also reported that within same *emm* type there were different *sof* types. They are *emm* 4 (*sof4* and *sof2920*), *emm* 25(*sof25, sof4958, sof75*), *emm*68 (*sof68, sof4470, sof4438*), *emm81* (*sof81, sof1658, sof1965*), *emm*88 (*sof88* and *sof1482*), *emm*89 (*sof89, sof4835*), st4935 (*sof4935, sof1881, sof27G*) and st448 (*sof448* and *sof3894*). Beall also reported that within different *emm* type there are same *sof* types. They are *sof2967* (st2967 and st1160), *sof* 90(emm 90 and st833), *sof8* (*emm* 8 and st3018) and *sof27G* (*emm* 27G and st4935). This relationship

between different *sof* types and same *emm* type and vice versa indicates inter strain recombination [223].

A study from Italy highlighted that the most prevalent *emm* type in throat population were *emm*12, *emm* 22, *emm* 4, *emm* 1, *emm* 89 and *emm* 5 [224]. These *emm* types were found to be macrolide resistant and carried *speC* gene, while the *speA* restricted *emm* types *emm*1 and *emm*3 were found to be macrolide sensitive [225]. Antibiotic resistance to invasive strains was mostly related to *emm* 89 [226]. The different incidence of exotoxin and antibiotic resistance genes among population did not appear to have intrinsic clinical significance, but may reflect the propensity of these traits to be associated with certain *emm* types independent of the source from which the strains were isolated [227].

emm typing has served as an epidemiological tool and the knowledge gained from various studies provided the impetus for exploring *emm* based genotyping system. The epidemiological studies from United States, Europe and the Caribbean reported *emm* types associated with pharyngitis, they are M1, M3, M5, M12, M18, M19 and M24. These M types are rarely found in impetigo lesions. These finding led to widely recognized concept that there are distinct GAS population of "skin types" and "throat types" [228].

In 1992, the distribution of GAS M types among strain isolated during 1988 to 1990 from patients with severe systemic infection from all regions of the United States were compared with the distribution of community GAS isolates. The study indicated that M1 and M3 were significantly more frequently associated with invasive GAS infections than with uncomplicated pharyngeal infections [99]. Another epidemiology study from United States

during the period 1995-1999 has also found *emm* types 1, 28, 12, 3 and 11. This accounted for 49.2% of isolates [229].

In accordance with other studies M1, M3, M12 and M28 isolates were most frequently present [229, 230]. Interestingly, the different manifestations of invasive GAS disease (meningitis, arthritis, necrotizing fasciitis and puerperal sepsis) were found to be associated with certain M types. M2 was found exclusively in puerperal sepsis, M6 was associated with meningitis and M12 was associated with arthritis. M1 was found equally distributed among different manifestation of invasive GAS disease [231]. M1, M3, M6, M18 and M24 are considered rheumatogenic, while M1, M12, M55 and M57 are nephritogenic [1](REF).A Study from North India has shown that out of 94 GAS isolate analyzed thirty seven *emm* types were identified and there was no significant association found between *emm* type and the source of isolation [232].

Few reports examining *emm* type diversity has been reported, however most report support the presence of predominant *emm* types. A community study in Korea found that *emm*78 and *emm*23 accounted for 69% of GAS isolates in one region, whereas in another region M5, M12, M3 and M23 accounted for 52% of isolates [233]. M1, M3 and M6 accounted for 67% of isolates among circulating *emm* types from Mexico among patients with pharyngitis [234]. Nguyen et al. found that M1 and M2 clones represented 70% of GAS isolates collected in a survey of 52 patients with pharyngitis [235]. By contrast studies with isolates collected from Northen Teritory of Australia revealed much greater diversity of GAS strains [236] . Study from North India revealed 33 *emm* types out of 59 GAS isolates, few were new types and no predominant clone was found [124].The factors that influence the epidemiological distribution of particular *emm* types are unknown.

Certain studies have pointed to an association between particular *emm* types and severe GAS infection [237, 238]. A number of studies have also suggested that serious GAS disease is associated with emergence of highly virulent GAS clones [239, 240] whereas others have found no evidence of this association [241-243]. The concept of M types with increased virulence capacity therefore remains controversial [244].

3.7.1.2. *emm* pattern typing

Another approach known as *emm* patterning is a method of classification that recognizes differences in the number, arrangement and subfamily content of *emm* gene in the GAS chromosome [137]. Five patterns designated A to E are currently in use. An *emm*-pattern A-C gene arrangement is generally considered to be associated with isolates from throat infections, while those from *emm* pattern D are related to skin infections. Isolates containing the *emm* pattern E arrangements of *emm* genes display an affinity for both tissues [137]. A large number of isolates from many geographic locations have been related to these associations [245].

The *emm* pattern is a genetic marker that distinguishes many throat and skin strains of GAS [245]. The *emm* pattern is defined by the chromosomal arrangements of *emm* subfamily genes. Plasminogen binding M protein encoded by *emm* gene is found to be uniquely associated with *emm* pattern D [246]. A study from Ethiopia has shown that there is strong association between strains isolated from throat and *emm* pattern A-C, but failed to show the association for the strains isolated from skin [247].

3.7.1.3. Vir- typing

Vir typing, another molecular typing method based on the restriction fragment length polymorphism (RFLP) of a PCR based amplified region of the 4-7 kb Vir regulon of GAS [248]. This locus encodes M and M related proteins and their positive regulator (the *mga* product)[249]. This method has been found to be discriminatory and is applicable to all isolates from diverse geographic origins [236].This method came into use, when there was more M non-typeable GAS isolates were identified from Northern Teritory of Australia.

In a comparative study involving Vir typing, multi locus enzyme electrophoresis (MLEE), random amplification of polymorphic DNA (RAPD) and serological techniques, Vir typing was found to produce unambiquous and easily scorable restriction fragment length polymorphism (RFLP) patterns, easier to perform and interpret than other molecular typing techniques.Some VTs containisolates from more than one M type. Of the 15 isolates which belong to VT16, five were serotyped as M1, nine were serotyped as M55 and one was serotyped as M80 [250].

Multiple typing methods have been used in the characterization including M and T serotyping, antibiotic resistance typing, ribotyping, random amplification of polymorphic DNA, pulse field gel electrophoresis, restriction fragment length polymorphism, vir typing, *emm* sequence typing and multi locus sequence typing. Although all these methods have proven useful for the characterization of GAS isolates, phenotyping methods have declined in popularity and the mainstream methods are now *emm* sequence typing, pulse field gel electrophoresis and multi locus sequence typing.

3.7.1.4. Multi locus sequence typing

Multi locus sequence typing (MLST) is a nucleotide based method that is well suited towards characterization of genetic relationship between the organisms of a bacterial species. Housekeeping genes are chosen for analysis because they are present in every organism and mutations within them are very neutral. Clones, defined as isolates that are descendants of a recent common ancestor, can be identified as having shared alleles at each of the housekeeping loci [251].

For MLST of GAS seven housekeeping loci's are used. MLST is a tool that can be used to study the molecular epidemiology and population genetic structure of microorganisms. MLST of 212 GAS isolates with 78 different *emm* types identified hundred unique combinations of allelic profiles [38].

Studies from Australia has revealed , most of the *emm* types observed among GAS isolates from an aboriginal island community were found in other parts of world, and few *emm* types were recovered within island community. The multi locus genotypes of isolates of the same emm type from different regions were different .This findings has implications for attempts to make global associations between *emm* type and certain disease manifestations [252].

Population genetic structure and phylogenetics are powerful tools. The site specific mutagenesis at codon under diversifying selection provides a rational approach for studying the effect of each adaptive change on the in vitro functional activity and immunogenicity of streptokinase. Isogenic mutants, generated by direct allelic replacement of the parental *ska* gene with an *ska* allele of another lineage can be used to measure biological properties of GAS [246].

Two hundred and twelve GAS strains collected from 34 countries was used to characterize MRGAS (Macrolide Resistant Group A Streptococci). Twenty two MRGAS clonal complexes were recovered from more than one continent; intercontinental strain represented 80% of the MRGAS of the study. The findings suggest that horizontal transfer of macrolide resistant clones and their descendants are both major components of the macrolide resistance problem found within species [36].

MLST for GAS has been reported from other parts of the world (Australia [252], United States [38], London [253], Germany [254], Nepal [125], and Norway[255]), but there are no reports from India, studying the population genetic structure of GAS.

3.8. Treatment

Empirical treatment of clinical pharyngitis using penicillins is a common in general practice. In case of penicillin allergy erythromycin or cephalexin is the drug of choice. Penicillin cannot be recommended in highly endemic situations, for several reasons. Firstly, oral antibiotics do not give as much coverage as benzathine penicillin, to patients who may later go on to develop RF/RHD due to lack of adequate antibiotic protection. Secondly, misuse of penicillin may pave way to increased penicillin tolerance. Thus scientifically, it is better to institute the treatment after confirming a GAS etiology which would reduce the unnecessary use of antibiotics for treatment of pharyngitis[256].

Erythromycin resistance has been reported as high as 70% GAS isolates in Japan and nearly 50% of pharyngeal isolates in Finland. In Southern India during 1986-2002 the overall erythromycin resistance was 2.7%. All isolates were susceptible to penicillin[221].

Treatment of rheumatic fever has three goals: eliminating any residual streptococcal infection; reducing inflammation, particularly in the joints and heart, and thus relieving symptoms; and limiting physical activity that might aggravate the inflamed structures. IN India the following guidelines are recommended for the management of rheumatic fever and rheumatic heart disease, they are as follows, (1) Streptococcal eradication with appropriate antibiotics (Benzathine penicillin single dose or penicillin V oral or azithromycin). (2) Diagnosis of rheumatic fever based on Jones criteria. (3) Control inflammatory process with aspirin with or without steroids (total duration of treatment of 12 weeks). (4) Treatment of chorea according to severity (therapy to continue for 2-3 weeks after clinical improvement). (5) Protocol for managing cardiac complication like valvular heart disease, congestive heart failure and atrial fibrillation. (6) Secondary prophylaxis with benzathine penicillin and management of anaphylaxis .

3.9. Vaccines

Research towards development of safe and efficacious vaccine against GAS infection is an ongoing and would be of major importance, especially in developing countries where the prevalence of rheumatic fever and post streptococcal glomerulonephritis are high causing a significant impact on public health. Different vaccine approaches include the M protein, C5apeptidase, cysteine protease, 26- multivalent vaccine, conserved C terminal region of M protein, putative proteins and surface markers such as fibronectin binding protein, group A carbohydrate protein antigen. They have been associated with reduced colonization and in some cases evoked protective immune response when tested in animal models.

Most GAS vaccine strategies have focused on the surface M protein, a major virulence factor of GAS. Antibodies to the amino terminus of the M protein have been shown to be

opsonic and provide protection against challenge from homologous organisms, which indicate they are specific. Some of the M protein based vaccines are PepM24 vaccine, recombinant M5 protein fragment, passive administration of M protein-specific IgA antibody, conserved epitopes found in the carboxy terminal region of M protein, conserved region[257], protein peptides conjucated to cholera toxin B subunit[258], carboxy terminal region of M protein [259] in vaccinia virus and carboxy terminal region of M protein in *Streptococcus gordoni* [260].

However, vaccine based on the M proteins face major problems. Since streptococcal sequelae are autoimmune type of disorders, molecular mimicry may play an important role in their pathogenesis, safety issues are major consideration in development of GAS vaccine. There are various drawbacks of the M protein based vaccines [261], namely, the amino terminus of the M protein is highly antigenically variable, there are more than 100 reference GAS *emm* types and an increasing number of which are not M serotypeable, large number of different M types circulating and by the dynamics in the epidemiology at a given geographic area, cross reaction with human tissues resulting in antibody response that cross react with human myosin. Moreover proper animal models and large scale clinical trials involving thousands of participants including children will be needed to assess both the safety and efficacy of GAS vaccines.

The *emm* types of GAS, which causes invasive disease in the United States, is promising in terms of use of the 26-valent vaccine but there may be limitations with regard to global administration of this vaccine [262]. There is need for a GAS vaccine in both the developing and developed country, but the needs for the region differs. Understanding these needs and

difference in the epidemiology, as well as the molecular basis of GAS virulence, should promote further development.

4. SCOPE AND PLAN OF WORK

As submitted to the university along with the application for the PhD programme, shown here as a university requirement in the thesis.

4.1. Study Design

Phase I: *emm* typing of GAS isolated from the community (school survey) and from clinical cases (hospital isolates).

Phase II: MLST of GAS from the community and clinical samples.

Phase III: Determination of phylogeny of GAS isolates by comparing dendrograms based on *emm* types and STs

Phase IV: Clonality and allelic variations among GAS isolates by eBURST analysis

Phase V: Determination of linkage disequilibrium among GAS isolates based on their STs.

4.2. Research Questions

- Do *emm* types of different groups of community GAS isolates differ from the clinical isolates in their distribution?
- What are the MLST profiles of GAS isolates causing human infections and circulating in the community?
- Do such patterns differ among community isolates and those causing invasive diseases?
- Can the genetic heterogeneity of GAS isolates be better studied by genetic variations of the seven housekeeping genes?
- Can MLST explain the variations seen among the GAS isolates circulating in an endemic community?

5. MATERIALS AND METHOD

5.1. Sample collection

GAS (GAS) isolates selected for this study included two groups:

a) Those recovered from rural school children aged 5-11 years during an ongoing epidemiological project funded by Indian Council of Medical Research. These will be referred to as community isolates.

b) Those recovered from patients seen in the outpatient departments of a tertiary care hospital of the Christian Medical College (CMCH). These will be referred to as clinical isolates

5.1.1 Community isolates

These were recovered from school children aged 5 to 11 years and attending a rural school situated 12 Kms from Vellore town. The GAS isolates included those recovered from children with clinical pharyngitis (Community Pharyngitis Isolates), those colonizing throat of normal healthy children (Normal Throat Isolates), those recovered from clinical impetigo (Impetigo Skin Isolates) and those colonizing normal skin (Skin Isolates). The isolates were identified as GAS by micro-nitrous acid extraction method, a co-agglutination technique standardized in our laboratory and using in-house prepared antisera. The GAS isolates were preserved by lyophilization and stored at $+4^{o}$C until testing.

5.1.2 Clinical isolates

These included GAS isolates consisted of two groups.

a) Those recovered from throat cultures of patients with acute upper respiratory tract infections (Hospital Pharyngitis Isolates), seen at CMCH during January 2002-October 2004

b) Those recovered from patients with invasive GAS disease seen at CMCH. (Invasive isolates). Invasive isolates were defined as those isolates recovered from normally sterile body fluids and were isolated in our laboratory from January 2000- July 2007. All isolates were preserved in lyophilized vials and stored at $+4^{o}$C until testing. The clinical details of patients with invasive disease and acute pharyngitis were collected from their hospital charts. Their *emm* types were already available for comparison.

5.2. *emm* typing of GAS isolates

This was carried out by the technique devised by the Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, USA and standardized in our Laboratory [32]. The method was as follows:

5.2.1. DNA Extraction

- With a loop fair amount of fresh growth was picked up, emulsified in 300 µl of 0.85% Sodium chloride and heated at 70°C for 15 minutes.
- The samples were spun at 13,000 rpm for 2 minutes.
- The pellet was re-suspended in 50 µl TE (10mM Tris, 1mM EDTA, pH8) containing 300U/ml mutanolysin and 30µg/ml hyaluronidase (Sigma-Aldrich, Bangalore)
- Incubated 37°C for 30 minutes and heated at 100°C for 10 minutes.
- The lysates were Stored at -20°C until use.

5.2.2. DNA Amplification

The extracted DNA was amplified by PCR as follows.

A 20µl of master mix was prepared as follows (TAQ PCR CORE KIT, Qiagen, Hilden, Germany)

- Primer(forward) - 0.4 µl (TATT (C/G) GCTTAGAAAATTAA) (Sigma-Aldrich, Bangalore)
- Primer(reverse) - 0.4 µl (GCAAGTTCTTCAGCTTGTTT)
- 10X buffer - 2 µl containing 1.5mM $MgCl_2$
- d NTP - 0.4 µl
- Taq polymerase - 0.06 µl
- Water (milliQ) - 16.74µl

0.5µl of template was added to 19.5µl of master mix.

The thermal cycler conditions included initial denaturation at 94°C for 1minute followed by 94°C for 15 seconds, 46.5°C for 30 seconds and 72°C for 1 minute 15 seconds repeated for 10 cycles and followed by the same conditions but with 10 seconds increment for 72°C for each of the subsequent 19 cycles. The amplicon was maintained at + 4°C.

The *emm* amplicon was visualised by electrophoresis on a 1% agarose gel. Electrophoresis was carried out at 100 V for 50 minutes with an appropriate base pair marker. The quality and quantity of the amplified product was checked by visualization of the gels stained with 10µg/ml ethidium bromide in a gel documentation system (BIO-RAD, California USA). The single band in range of 700 to 1200 base pairs shows that the amplified product is specific and pure.

5.2.3. *emm amplicon* sequencing

5.2.3.1. Pres-sequencing clean up

The *emm* amplicons were subjected to pre-purification by Millipore filtration technique (Millipore Corporation, Bedford, MA, USA) to remove unused dNTPs and primers. The following steps were followed.

- The amplified product was made upto 100 µl using sterile distilled water.
- The solution was then transferred to Millipore micrtiter plate and attached to the Millipore vacum manifold.
- Vacum pressures was applied for approximately 10 minutes or until the well was dried.
- One hundred microliters of sterile distilled water was added and the same process was repeated.
- Twenty micrliters of sterile distilled water was added to the wells, mixed 25 times and the contents transferred back to PCR tubes.
- Electrophoresis was done with 1% agarose gel in 0.5X Tris-borate buffer with EDTA (TBE buffer) for pre-purified product to confirm purity of *emm* amplicon.

5.2.3.2. Sequencing Reaction

Prepare 10 µl reaction mix as follows,

- RR mix - 2µl
- Buffer - 2µl
- Primer - 1.6µl (TAT TCG CTT AGA AAA TTA AAA ACAG)
- Template - 2µl
- Distilled water - 2.4µl

The PCR parameters were as follows:

- 96 °C for 15 seconds
- 96 °C for 20 seconds
- 50 °C for 20 seconds
- 60 °C for 4 minutes

} 24 Cycles

- 72°C for 10 minutes

5.2.3.3. Post-sequencing clean up

The sequencing PCR products were then purified by membrane filtration technique kit (Millipore Corporation, Bedford, MA, USA).The following steps were carried out for post cycle-sequencing clean up procedure.

- The total volume of the amplicon was made up to 40 µl with injection solution and transferred to the Millipore micro titer plates in the vacuum manifold.
- Vacuum pressure was applied to the plates until the plates are dry.
- Forty micro liters of injection solution was added to the wells and the vacum process was repeated.
- Thirty micro liter injection solution was added, the contents were mixed well by pipetting 25 times and transferred to tubes for sequencing.

The sequencing was done with the Big Dye terminator Kit in an ABI PRISM 310 automated sequencer (Applied Biosystems, Warrington, UK).

5.2.4. Assigning *emm* types

The chromatogram and the text sequences obtained were edited and the sequences were blasted to assign a type and subtype and the type specific database with at least the first 240 bases obtained with primer *emm*seq2 (www.cdc.gov/ncidod/biotech/strep/strepblast.html).

More than 95% homology with type specific 150 bases of the standard reference sequence gives the designated *emm* type of the test strain. If homology is not 95%, it was compared with most similar *emm* types and if new, the sequence traces were submitted for confirmation. New types were identified by the curator of this site, Dr. Bernard Beall, at Centers for Disease Control and Prevention (CDC), Atlanta (bbeall@cdc.gov) on the basis of sharing less than 92% sequence identity over the first 90 bases encoding the deduced processed M protein of the *emm* type reference strain.

For the subtype assignment, database of trimmed 180 base entries corresponded to the first 50 residues of the mature M protein and the adjacent 10 C terminal residues of the signal sequence. If a perfect 180/180 match was obtained to an entry from the type-specific BLAST option, the subtype was reported to be correctly identified. If a perfect match to bases 31-180 is combined with 3 or fewer mismatches to bases 1-30 was found, this also indicated identification of the specific subtype. If there was any mutation in the DNA sequence corresponding to the first 50 residues of the mature M protein, it was considered as a new subtype.

5.3. Multi Locus Sequence Typing Protocol

Multi locus sequence typing is a nucleotide sequence method where in seven housekeeping loci are chosen for characterization of GAS [38]. They are internal fragments of the following enzymes whose primers (Sigma-Aldrich, Bangalore) are also given below.

1. Glucose kinase (*gki*)
    - *gki-up* GGCATTGGAATGGGATCACC
    - *gki-dn* TCTCCTGCTGCTGACAC

2. Glutamine transporter protein (*gtr*)

- *gtr-up* GAGGTTGTGGTGATTATTGG
- *gtr-dn* GCAAAGCCCATTTCATGAGTC

3. Glutamate racemase (*murI*)

- *murI-up* TGCTGACTCAAAATGTTAAAATGATTG
- *murI-dn* GATGATAATTCACCGTTAATGTCAAAATAG

4. DNA mismatch repair protein (*mutS*)

- *mutS-up* GAAGAGTCATCTAGTTTAGAATACGAT
- *mutS-dn* AGAGAGTTGTCACTTGCGCGTTTGATTGCT

5. Transketolase (*recP*)

- *recP-up* GCAAATTCTGGACACCCAGG
- *recP-dn* CTTTCACAAGGATATGTTGCC

6. Xanthine phosphoribosyl transferase (*xpt*)

- *xpt-up* TTACTTGAAGAACGCATCTTA
- *xpt-dn* ATGAGGTCACTTCAATGCCC

7. Acetyl-coA acetyl transferase (*Yqil*)

- *yiqL-up* TGCAACAGTATGGACTGACCAGAGAACAAGATGC
- *yiqL-dn* CAAGGTCTCGTGAAACCGCTAAAGCCTGAG

5.3.1. DNA Extraction for MLST GAS isolates :(QIAamp DNA BLOOD mini kit, Qiagen, Hilden Germany)

This was done using a commercial kit (QIAamp DNA BLOOD mini kit, Qiagen, Hilden Germany) and was briefly as follows:

- Inoculated the test GAS strain into 8 ml Todd-Hewitt broth in a 15 ml tube and incubated overnight at 37°C.
- Pelleted the broth culture by centrifugation
- 20mg/ml lysozyme was added to the enzymatic lysis buffer.
- Re-suspended the pellet in 180 µl of enzymatic lysis buffer and incubated for 60 minutes at 37°C
- Added 25 µl of proteinase K, 200 µl of buffer AL and vortexed for 5 seconds.
- Incubated at 70°C for 30 minutes.
- Added 200 µl of ethanol (95%) and vortexed for 5 seconds.
- Pipetted the mixture into a column provided in the kit (QIAamp spin columns) and centrifuged for 2 minutes at 10,000rpm; centrifugation was repeated to remove the liquid completely form the column.
- Discard the collection tube with the flow through.
- Place the column in a new collection tube; add 500 µl of buffer AW1 and centrifuged for 1minute at 10,000rpm and discarded the collection tube.
- Placed column in a new collection tube, added 500 µl of buffer AW2 and centrifuged for 3 minutes at 13,000 rpm.
- Discarded the collection tube.
- Placed the column in a 1.5ml micro-centrifuge tube, added 200 µl of buffer AE, waited for 1 minute and centrifuged for 1 minute at 10,000 rpm.
- Repeated the previous step and discarded the column.

5.3.2. DNA Amplification

The extracted DNA was amplified by PCR as given below.

50 µl of master mix was prepared using a commercial kit (TAQ PCR CORE KIT, Qiagen, Germany, Hilden) .

- Primer(forward) - 1 µl
- Primer(reverse) - 1 µl
- 10X buffer - 5 µl
- dNTP - 5 µl
- $Mgcl_2$ - 5 µl
- Taq polymerase - 0.46 µl
- Water (milliQ) - 32 µl

To 49 µl of master mix added 1 µl of template.

The PCR parameters were as follows

- 95°C for 5minutes
- 95°C for 1minute
- 55°C for 1minute
- 72°C for 1minute

} 28

- 72°C for 5minute

The PCR products were checked for amplification by electrophoresis on 1% agarose gel. Electrophoresis was carried out at 100 V for 45 minutes. The quality and quantity of the amplified products were checked by visualization of the 10µg/ml ethidium bromide stained bands in a gel documentation system (BIO-RAD, California, USA).

The amplified products were pre-purified and checked for purity by gel electrophoresis. The sequence of each fragment was obtained using the respective primers used in the initial amplification.

5.3.3. Assigning Sequence types (ST)

The forward and reverse sequences generated were compared aligned, analyzed, edited, and trimmed exactly to the respective allele. The size of the amplicon for different loci used for assigning alleles for Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*), Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*) are 498, 450 ,438 405, 459, 450 and 434 base pairs respectively.

The single locus query or multiple locus options at http://spyogenes.mlst.net determines an allele number for each sequence trimmed and pasted for all seven loci of a test strain into the corresponding boxes. The software checks for correct length of sequence for respective locus. The sequence of each fragments are compared with all the previously identified sequences (alleles) at that locus. If the sequence corresponds to the known allele, an allele number is given to the sequence. If the sequence is a new allele it should be compared with the most similar allele for that locus to check for nucleotide differences and if new, the sequence traces (forward and reverse) should be submitted to the database curator Mr Daniel Godoy**,** (d.godoy@imperial.ac.uk) Research Technician, Division of Epidemiology, Public Health and Primary Care, Imperial College, London. The curator will check the sequence trace, provide a new allele number and add the new allele to the database. The combination of the seven allele number forms the allelic profile of the strain and each different allelic

profile is assigned as a sequence type (ST).The MLST databases are hosted on two web servers located at Imperial College London (http://www.mlst.net) and Oxford University (http://pubmlst.org).

5.4. eBURST (Based upon related sequence types) Analysis

BURST was devised and developed by Ed Feil [263], to determine the relationship between closely related isolates of a bacterial species or population. BURST explains the evolutionary relationship among closely related genotypes in a simple manner and predicts the founding genotype of each group. It displays a radial diagram with the predicted founder centered. eBURST was developed to analyse the multi locus sequence typing data and integrated in the MLST website by David Aanensen. The MLST data used by eBURST are the STs and their associated allelic profiles.

The assignment of clonal complexes and clonal variants from the allelic profile can be carried out using an algorithm called BURST (Based upon related sequence types) (http://eburst.mlst.net/). eBURST uses the sequence types (STs) and their associated allelic profiles as input data. To identify groups of related STs, the most conservative definition where all members assigned the same group share identical alleles at ≥ 6 of the 7 loci with at least one another member of the group. A group is used as a neutral term for the collection of sequence types that are clubbed together by eBURST. A clonal complex is a set of STs that are descended from the same founding genotype (6/7 shared alleles) and differ from the founding genotype or predicted founder by a single loci. This approach results in non-overlapping groups that is no ST can be assigned to more than one group. The predicted primary founder of the group is defined as the ST that differs from the other STs by only a

single locus. If they differ from one another by single loci, they are single locus variant (SLV) and if they differ by two locus they are known as double locus variant (DLV).BURST does not make any inference about the relationship between distantly related genotypes that belong to different groups.

eBURST uses statistical support and bootstrapping procedure to provide the level of confidence in the assignment of primary founder in each group. The bootstrap values are only available using the stringent group definition where all members assigned the same group share identical alleles at ≥ 6 of the 7 loci. The bootstrap value shown for each ST is the percentage of times the ST was predicted to be the primary founder of the group. The bootstrap values for a ST may differ when one of the ST is used, compared to when all isolates are included, since the STs represented in each re-sampling will be influenced by the frequencies of the STs in the database. The population snapshot (whole MLST data) can be acquired by relaxing the definition from 6/7 to 0/7.This will arrange the whole input data into a single eBURST group and will display all the clusters of STs and unlinked individual STs as a single BURST diagram, this display is called "Population snapshot". The linked clusters within this population snap shot represents clonal complexes and the primary founders and the sub founders of these linked clusters are colored blue and yellow dots respectively. Each black dot represents an ST and the size of the dot reflects the number of GAS isolates in each ST.

5.5. Construction of phylogenetic tree based on *emm* types

The Phylogenetic tree was constructed by using MEGA-4 [264]. The evolutionary distances were computed using the Maximum Composite Likelihood method. 500 base pairs of representative 49 *emm* types subjected to multi locus sequence typing from the community and the hospital GAS isolates were used to construct the phylogenetic tree.

5.6. Construction of phylogenetic tree based on Sequence Types (ST).

Phylogenic tree based on the STs of MLST was constructed using the concatenated sequence of the ST. Concatenated sequence was obtained by fusing all the seven loci's from the allelic profile. The concatenated sequences formed were of 3,314 base pairs. The evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number .The tree was constructed by using MEGA4 [264, 265].

5.7. Linkage disequilibrium

The detection of significant level of linkage disequilibrium may indicate that recombination has not been sufficiently frequent to randomize the association between alleles. The level of linkage disequilibrium can be calculated from the distribution of allelic mismatches between all GAS isolates (the number of difference in the allelic profile).The observed variation in the mismatches of alleles, V_o is compared with the expected variance $V_{E.}$ A useful measure is the index of association (I_A)

$$\mathbf{I_A = (V_O / V_E) -1}$$

Observed variance (V_O)
Expected variance (V_E)

Index of association will not be significantly different from zero since $V_{O \text{ and}}$ V_E should be approximately the same in a population with linkage equilibrium.

The Linkage disequilibrium program (http://linux.mlst.net/link_dis/index.htm) analyses whether the alleles at different loci within a population of bacteria are randomly associated (in linkage equilibrium), or whether there is a significant association between the alleles at different loci (linkage disequilibrium) [266]. The sequence types of each strain and the allelic profiles is cut and pasted into the window provided. The observed variance in the distribution of allelic mismatches in all pair-wise comparisons of the allelic profiles is computed and is compared to that expected in a freely recombining population (linkage equilibrium). This is implemented by computing the maximum variance obtained using 20, 100 or 1000 randomizations of the data set. Significant linkage disequilibrium is established if the observed variance, obtained with the real data set, is greater than the maximum variance observed with any of the randomized data sets with p values of 0.05, 0.01 or 0.001 when the observed variance is greater than that found in any of 20, 100 or 1000 randomized data sets, respectively. If the observed variance is less than that obtained with the randomized data sets there is no evidence of a departure from linkage equilibrium.

5.8. Maximum nucleotide sequence identity

Maximum nucleotide sequence identity among allelic variants of diverse STs having the same *emm* type(s) were carried out by clustal W (http://www.ebi.ac.uk/Tools/clustalw/) [267].Sequence alignment is a way of arranging the sequences of DNA to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically

represented as rows within a matrix. Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns.

If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps insertion or deletion mutations introduced in one or both lineages in the time since they diverged from one another.

RESULTS

6.1. Selection of GAS isolates for *emm* typing

6.1.1. Community isolates

A total of 319 GAS isolates recovered from cases of pharyngitis *(n = 101)*, impetigo *(n = 65)*, throat cultures of normal healthy children *(n = 113)* and skin cultures of normal healthy children *(n = 40)* were selected for *emm* typing. These were recovered from rural school children during an epidemiological study conducted from 2000 – 2002.

6.1.2. Clinical isolates

Ninety Nine GAS isolates from clinical sources were subjected to *emm* typing. These included isolates from hospital diagnosed acute pharyngitis *(n=41)* and invasive isolates *(n=58)* that includes isolates from blood *(n=35)*, peritoneal fluid *(n=7)*, CSF *(n=5)*, pleural fluid *(n=5)*, fluid from necrotizing fasciitis *(n=2)* and cellulitis *(n=1)*, bile, synovial fluid and fluid from retained placenta *(n=1)* each.

6.2. Distribution of *emm* type among community GAS strains

Three hundred and nineteen GAS isolates from the school survey were subjected to *emm* typing **(Table.1).** A total of 73 *emm* types were identified among them, of which 17 were sequence types and two were hitherto unrecognized types. Sixteen types accounted for 166 GAS isolates (52%) and these included 112, 11, 82, 100, 63, 93, 85, 105, 108, 49.4, 110, 113, 1, 44, 66 and 111.

The most predominant *emm* type was *emm* type 112 which accounted for 29 (9 %) of the GAS isolates and was found predominantly in normal throat *(n=18)* and impetigo skin *(n= 8)*

6.2.1. Distribution of *emm* types among Community associated Pharyngitis GAS isolates

Forty five *emm* types were identified among 101 community pharyngitis GAS isolates **(Fig 1)**. Twelve *emm* types namely 11, 110, 108, 63, 85,1,55,58,74,18 and 92 accounted for 53 (52.4%) of the 101 isolates. Type 11 was the most predominant *(n=9)* followed by type 110 *(n=8)* which was found only among pharyngitis isolates. Type 112, the most predominant type among all community GAS isolates, accounted only for 0.9% from the community associated pharyngitis. Types 1 and 3, the commonest invasive types accounted for only 4 isolates (3.96%).

6.2.2. Distribution of *emm* types among normal throat carriers

Among 113 GAS isolates from normal throat, 47 *emm* types were identified **(Fig. 2)**. Predominant types including 112 *(n=18)*, 66, 105*(n=5 each)* and 11, 49.4, 82, 113 and 111*(n=4 each)* accounted for 48 (42%) of the 113 Isolates. Type 112 *(n=18)* was the most predominant *emm* type while two new types, stKNB3 *(n=2)* and stKNB8 *(n=1)* are two new *emm* types accounted for 2.6% of GAS isolates.

6.2.3. Distribution of *emm* types among GAS isolates causing impetigo

Among 65 GAS isolates causing impetigo, there were 35 *emm* types out of which, three types namely 112, 100, 93 and 44 accounted for 21 (32.3%) of 65 isolates **(Fig. 3)** . Type

112 was the most predominant type accounting for 12% of the GAS isolates while type 1 accounted only for 1.5%.

6.2.4. Distribution of *emm* type among normal skin

Twenty five *emm* types were identified among forty GAS isolates.The predominant *emm* types colonizing skin included 93 *(n=4),* 36 and 44 *(n=3 each)* which together accounted for 25% of the 40 isolates **(Fig. 4)**.

6.3. Distribution of *emm* types among Hospital GAS isolates

Ninety nine non repetitive GAS isolates from diverse clinical specimens were subjected to *emm* typing. These included isolates from cases of acute pharyngitis *(n=41)* and invasive isolates *(n=58).* Invasive GAS isolates were recovered from cases of sepsis *(n=13),* meningitis *(n=9),* peritonitis *(n=8)* ,cellulitis *(n=5)* , pneumonia *(n=5),* necrotizing fasciitis *(n=3)* and umbilical infection*(n=3),* arthritis*(n=2)* ,puerperal sepsis*(n=1)* ,retained placenta*(n=1)* ,empyema *(n=1)* ,endocarditis*(n=1),*gangrene*(n=1*) ,pleural effusion*(n=1),* septic arthritis*(n=1)* ,urogenital*(n=1).*

6.3.1. Distribution of *emm* types among GAS causing acute pharyngitis in patients seen in the hospital (Hospital Pharyngitis)

Thirty *emm* types were identified among 41 GAS isolates from cases of hospital pharyngitis. Types 75 *(n=4),* 77 and 55 *(n=3each)* accounted for 10 (24.3%) of 41 isolates **(Fig. 5)**.

6.3.2. Distribution of *emm* types among invasive GAS isolates

A total of 37 *emm* types were identified among 58 invasive GAS isolates **(Fig 6)**. Five types, namely, 110 *(n=4),* 63, 74, 85 and st854.1 *(n=3each)* accounted for 16 (27.5%) of 58 isolates Two new *emm* types, stKNB6 and stKNB9 *(n = 1 each)* were identified from invasive GAS isolates. Types 1and 3 were conspicuously absent.

6.3.3. Comparison of *emm* types associated with community and Hospital pharyngitis

Table 2 gives the distribution of 12 most predominant *emm* types of 101 community pharyngitis GAS isolates among hospital pharyngitis GAS isolates. Five types namely, 11, 63, 85, 1 and 55 were identified among 41 hospital pharyngitis isolates, accounting for 7 (17.1%) of the isolates. Type 55 was the only type that was seen among both groups as a predominant type.

6.3.4. Distribution of Invasive *emm* types among Community and Hospital Pharyngeal GAS isolates

The distribution of predominant *emm* types associated with invasive diseases among community and hospital pharyngitis GAS isolates is given in **Table 3**. The 15 *emm* types that accounted for 62% of the invasive isolates, 9 types were found among community isolates while only 3 types were found among hospital isolates. Types 63, 85 and 93were identified among all three groups.

6.4. Multi locus sequence typing (MLST) of GAS isolates.

A total of one hundred and forty three GAS isolates were subjected to multi locus sequence typing. Out of the 143 GAS isolates, seventy three GAS isolates were selected from the dominant *emm* types from the community which includes cases of pharyngitis, normal throat, impetigo and normal skin. The rest of the seventy GAS isolates selected for multi locus sequence typing were identified and isolated from hospital. This included twenty GAS isolates from hospital pharyngitis and fifty GAS isolates from invasive cases and sterile body fluids.

6.4.1. Community Pharyngitis GAS isolates selected for MLST

Seventy three GAS isolates representing 23 predominant *emm* types identified among 319 GAS isolates were subjected to MLST **(Table 4)**. These included pharyngeal isolates *(n = 21),* normal throat isolates *(n = 22),* impetigo isolates *(n = 18)* and normal skin isolates *(n = 12).* These 23 types accounted for 214 (67.1%) of the 319 isolates.

6.4.2. Hospital Pharyngitis GAS isolates selected for MLST

A total of 20 pharyngeal GAS isolates representing 18 predominant *emm* types seen among 20 pharyngeal isolates were selected for MLST.

6.4.3. Invasive GAS isolates for MLST

Fifty invasive GAS isolates including those blood *(n = 32)* and fluids *(n=18)* were subjected to MLST. The latter included CSF *(n=5),* Peritoneal fluid *(n=6),* Pleural fluid *(n= 4),*

Synovial fluid *(n=1)*, Bile *(n=1)* and aspirate from necrotizing fasciitis *(n = 1)*. These isolates accounted for 31 *emm* types.

6.5. Distribution of Sequence Types (STs) among GAS isolates

Table 5 gives the distribution of STs among 143 GAS isolates that were subjected to MLST. The sequence types are arranged chronologically. A total 93 STs were identified among them; of thes25 were known sequence types and 68 (73.1%) were new sequence types **(Table 6)**. The new alleles and allelic profiles were submitted to the Curator Mr Daniel Godoy**,** (d.godoy@imperial.ac.uk) Research Technician, Division of Epidemiology, Public Health and Primary Care, Imperial College, London who designated them as new types.

6.5.1. Distribution of STs among Community Pharyngeal GAS Isolates

Twenty STs were identified among 21 pharyngeal GAS isolates **(Table. 7)**; fifteen of these sequence types were new types. ST 499 was seen in two isolates of *emm* type 113.

6.5.2. Distribution of STs among Community Normal Throat Isolates

Twenty two STs were identified among 22 GAS isolates indicating that all isolates were genetically different **(Table.8).** Fourteen of the 20 types were new types.

6.5.3. Genetic relatedness and diversity among *emm* types of GAS isolates from cases of pharyngitis and normal throat from the community

Seven *emm* types identified among pharyngitis and normal throat GAS isolates had the same MLST sequences **(Table .9)**; however 8 same *emm* types found in the two groups had different ST **(Table .10)**.

6.5.4 Distribution of STs among Impetigo Isolates

Seventeen STs were identified among 18 GAS isolates from impetigo. Only one type, ST 100, was identified in two different *emm* types, ie.55 and 92. **(Table.11).** Seven types were new types.

6.5.5 Genetic relatedness and diversity among *emm* types of GAS isolates from cases of Impetigo and normal throat

Seven *emm* types found in both groups had the same ST while 8 types seen in both groups had different STs. **(Tables 12 & 13).**

6.5.6. Distribution of STs among Normal Skin Isolates

All 12 GAS isolates from normal skin belonged to 12 different STs **(Table.14)**. Eight of them were new sequence types. Four *emm* types had the same ST as that identified in impetigo isolates while 5 had different STs.

6.6. Distribution of Sequence types for Hospital Pharyngitis GAS isolates

A total of 19 STs were identified among 20 GAS isolates **(Table 15)**. Two *emm* types, 1 and 12 had the same ST 36. Eight new allelic profiles were identified which were assigned new sequence types, 471,479,482,483,485,501,508,510.

6.8. Distribution of STs among Invasive GAS Isolates

Forty-five ST types were identified among 50 invasive GAS isolates **(Tables 16 & 17)**. There were 31 STs among 32 blood isolates representing 25 *emm* types **(Table 16)**. Two blood isolates with *emm* type 85 had the same ST 484. On the other hand, three blood

isolates of *emm* type 110 had three different STs namely 522, 493 and 494. Similarly two blood isolates of *emm* type 74 had two different STs, 480 and 520. Thus, multiple isolates of *emm* types 74, 85, 86, 100, 110, and st6735 had all different STs.

Among isolates from other various body fluids, 17 STs were identified from as many *emm* types **(Table 17)**. Nine of them were new types. Two peritoneal fluid isolates with the same *emm* type, st854.1 had the same ST, 224. Moreover, one CSF isolate with *emm* type 74 had a ST 480 while one peritoneal fluid isolate with the same *emm* type also had the ST.

6.7. Comparison of ST's of same *emm* types identified among community and hospital pharyngitis GAS isolates

Of 26 *emm* types from hospital and community pharyngitis GAS isolates only few had the same sequence types. These were *emm* types 55, 77, 93, stNS554 and 63 with sequence types 100, 482, 10,510and 338 respectively **(Table.18).**

6.8. Distribution of different *emm* types with same ST

Of the ninety three STs identified in this study, four were shared by more than one *emm* type **(Table.19)**. Such STs are known as *emm* variable sequence types. There was no association between the STs and the sites from where the GAS was isolated **(Table 20)**.

Pair wise comparison of *emm* types sharing the same STs were done using clustal W (http://www.ebi.ac.uk/Tools/clustalw/). The maximum nucleotide identity was identified as follows,(ST 36) *emm* type 1 and 12 had a percentage identity of 50.47%, (ST100) *emm*55 and 92 had percentage identity of 53.42%, (ST 357) had % identity of 56.73% and (ST477) *emm* 44and 113 had percentage identity of 62.96%.

Twenty five of the 49 *emm* types subjected to MLST had more than one identifiable STs **(Table 21)** .Thirteen of them had >2 identifiable STs; of these, *emm* type 110 and 49.4 had five ST's identified among them. No association was found between the *emm* types and their STs with their site of isolation **(Table.21)**.

6.10 Allelic variations of different sequence types within same *emm* type

6.10.1. Allelic variations in *emm* type 1

Type 1 with four ST's showed allelic variations with respect to all seven housekeeping genes **(Table 22).** All were multi locus variants.

6.10.2. Allelic variations in *emm* type 12

emm type 12 had three different allelic profiles **(Table.23)** , they are ST 28,36,472. ST 36 which was found in *emm* type 1was also found in *emm* type 12.ST 28's allelic profile was entirely different from ST 36 and 472. The allelic profile ST 472 and 36 differs at the loci *gk*i and *mutS.*

6.10.3. Allelic variations in *emm* type 18

Type 18 with four ST's showed both single locus variant (SLV) and double locus variant (DLV). The STs 523, 526 and 535 were all SLV's with respect to *gtr* or *gki* while ST 120 showed variations with respect to five loci **(Table 24)**.

6.10.4. Allelic variations in *emm* type 44

ST 178,351,476and 477 were the sequence types of *emm* type 44 **(Table.25)** .The allelic profile of ST 477 showed variation with rest of the STs at all the loci except *xpt.* The

allelic profiles of ST 178 and 351 differed at *gki*, while the allelic profile of ST 178 and 476 differed only at *Yqil.*

6.10.5. Allelic variations in *emm* type 49.4

This type had five ST's identified among the various isolates from six different sources **(Table 26)**. ST 228 and 478 were single locus variants with respect to *gtr*; STs 529, 531 and 534 were also single locus variants between them, with respect to *gtr*; but they were double locus variant with STs 228 and 478 with respect to *gtr* and *muts*.

6.10.6. Allelic variations in *emm* type 85

The allelic profile of STs of *emm* type 85 **(Table.27)** showed variation at the loci *gki, gtr* and *yqil.* The STs 484,514 and 525 in comparison with ST 109 are single locus variant. ST 484 differs from allelic profile of ST 109 at the locus *gtr.* The other two STs 514 and 525 differ from ST 109 at the locus yqil and gki respectively.

6.10.7. Allelic variations in *emm* type 92

emm type 92 **(Table.28)** has three different STs , they are ST 100, ST485 and ST 524. ST 100 has an entirely different allelic profile compared to ST 485 and ST 524. ST 485 has an allelic variation at *gki* and *mutS* when compared with the allelic profile of ST 524.

6.10.8. Allelic variations in *emm* type 100

The STs of *emm* type 100 **(Table.29)** differs at the loci *gtr, recp* and *gki.* ST 486 and 487 are single locus variants of ST100 and differ at the loci *gki.* The ST 488 differs from ST 100 at the loci *gki* and *recp*, ST 488 is a double locus variant (DLV) of ST 100.

6.10.9. Allelic variations in *emm* type 105

emm type 105 **(Table.30)** had four allelic profiles which varied from one another at the loci *gki, gtr* and *xpt* . ST 151 differs from ST 490, ST 491 and ST 492 at the locus *gtr ,xpt* and *gki* respectively. ST 490,491 and 492 were single locus variants (SLV) of ST 151.

6.10.10. Allelic variations in *emm* type 108

ST 357,527and 533 were the STs found among *emm* type 108 **(Table.31)** .The allelic profile of ST 357 differed from other STs at all loci, Whereas the ST 527 and 533 showed variation at *murI* and *mutS.*

6.10.11. Allelic variations in *emm* type 110

The allelic profiles of different sequence types within the *same emm* type were analyzed to determine the extent of variations in the loci with respect to their source of isolation **(Table 32)**. The *emm* type 110 isolated from blood showed single locus variations (SLV). Type 110 with ST 493 and 494, both isolated from blood cultures, was different only with respect to one locus, ie. *gtr*. Thus these two were SLV's. The third blood culture isolate with ST 522 was also a SLV with respect to the other two blood culture isolates, but with respect to the locus, *yqil.* Type 110 from community pharyngitis with ST 495, and the bile isolate with ST 515 were also SLVs with respect to *gtr*, but a double locus variant with respect to the blood culture isolates.

6.10.12 Allelic variations in *emm* type 113

emm type 113 had three sequence types ,they were ST 477,499 and 500 **(Table.33)** . ST477 and 499 had allelic variation at the locus *yqil*, while ST 500 differs from ST 477at the loci *recp.*

6.11. Maximum nucleotide identity of same *emm* types with different allelic profiles

The different allelic profiles within same *emm* type were checked for Maximum nucleotide identity using Clustal W(http://www.ebi.ac.uk/Tools/clustalw/). The *emm* types having more than one sequence type were *emm*1 (ST15, ST36, ST469, ST530), *emm*12 (ST28, ST36, ST472), *emm*18 (ST120, ST523, ST526, ST535), *emm*44 (ST178, ST351, ST476, ST477), *emm*49.4 (ST228, ST478, ST529, ST531, ST534), *emm*85 (ST109, ST484, ST514, ST525), *emm*92 (ST100, ST485, ST524), *emm*100 (ST119, ST486, ST487, ST488), *emm*105 (ST151, ST490, ST491, ST492), *emm*108 (ST357, ST527, ST533), *emm*110 (ST493, ST494, ST515, ST522), *emm*113 (ST477, ST499, ST500). All the *emm* types showed a maximum nucleotide identity of 99% **(Table 34 – 45)** within their sequence types except *emm*108 which showed 98%. *emm* 108 showed 98% **(Table 43)** nucleotide identity with ST527 & ST533 with respect to ST357 whereas ST527 & ST533 had nucleotide identity of 99%. The sequence types within the *emm* types which showed 99% similarity varied from each other by 5 – 9 nucleotides. **Figure 7** shows the multiple sequence alignment of the different sequence types of *emm*1.

6.12. Population snapshot by eBURST of GAS isolates

The BURST (Based Upon Related Sequence Types**)** identifies mutually exclusive groups of related genotypes in the population and identifies the founding genotype of each group. **Figure 8** gives the eBURST population snap shot of 93 sequence types from 143 GAS isolates, for which MLST was carried out. The most stringent definition of an eBURST group is that all ST's assigned in the same group must share alleles at least six of the seven house- keeping loci with at least one other in ST in the group. This identified clusters of

closely related genotypes that are considered to have descended from the same founder and are defined as a clonal complex. Based on this definition, there were 19 groups with eight groups having predicted founders and the remaining groups did not have predicted founders. There were 41 singletons. In the **figure 8** 'the groups' are represented by black lines connecting the sequence types within groups. Out of the eight groups two groups had multiple sequence type as predicted founder within the group while the other six groups had single predicted founder, they are 534, 109, 120, 477, 119 and 535. The snap shot of BURST shows two clonal complexes with predicted founder ST 109 and ST 534 as the founder supported by bootstrap value showing more than 60% and is highlighted in blue dots.

Group 1 clonal complex had ST 534 **(fig.9a)** as predicted founder with a bootstrap value of 70% represented as a blue dot. The clonal complex consisted of ST 534 ST 228, ST517, ST531, ST529 and ST478. ST 534 had 4 single locus variant (SLV) and 1 double locus variant (DLV).The SLVs are represented as a pink line and DLVs are represented as blue line in the **figure 9a**.The four SLVs were ST 228, ST517, ST531 and ST529. ST 478 was the DLV of ST 534. In group 1 ST 534 the predicted founder had 4SLVs, while ST228, ST517, ST531, ST529 and ST478 had 3, 2, 2,2and 1SLVs respectively within the group. The STs of Group 1 clonal complex belongs to *emm* type 49.4 except ST 517 which belongs to stKNB6 which is a newly identified *emm* type.

Group 2 clonal complex **(fig.9b)** had ST 109 as predicted founder with bootstrap value 61%.The clonal complex was made up of five sequence types, they were ST 109, ST525, ST484, ST514 and ST528.ST 109 had three SLV and one DLV. The SLVs of ST 109 were ST 525, ST484 and ST514. In group 2 the ST 109 had 3SLV, ST525-2SLV, ST484, 514,528

had one SLV each. The STs in clonal complex with predicted founder 109 belongs to *emm* type 85 except ST528 which belongs to *emm* 55.

The group 3 clonal complex had multiple candidates as predicted founder and had bootstrap value 33%. The complex consisted of ST151, ST491, ST492 and ST490. The STs in this complex belongs to *emm* type 105. The rest of the groups had bootstrap value less than 50%.Group 4 clonal complex had ST120 as predicted founder with ST480 and ST520. The STs in this group belongs to *emm* 74 except ST120 which belongs to *emm* 18.

ST477, ST499and ST500 forms group 5 clonal complex with ST 477 as predicted founder .ST 477 falls under *emm* type 44 while ST499 and ST 500 belongs to *emm* 113.The sixth clonal group comprised of ST119, ST487 and ST486 and belong to *emm* 100 with predicted founder as 119. The seventh group of clonal complex had ST 535,526 and ST523 with the predicted founder ST535and belong to *emm* 18.The group eight had three sequence types and has multiple predicted founder.ST494, ST493 and ST522 forms the clonal complex of group 8 and belong to *emm* 110. All of these STs within group 8 differ from each other by SLV.

Figure 10 describes the single locus variants (SLV) found among the 93 sequence types of 143 GAS isolates. The SLVs are found only in the 19 groups created using the standard eBurst definition. SLVs are those which differ at one locus. The pink line indicates the SLVs found within the 19 groups. All the STs present in the 19 groups are SLV. There are singletons, which do not form the clonal groups in which the SLVs are absent.

Figure 11 depicts the double locus variant in the population snapshot of 143 GAS isolates with 93 sequence types. The sequence types which differ at two loci's are known as double

locus variant (DLV), they are represented as blue line in the eBURST population snap shot. The DLVs are found only in few groups. All the sequence types found in the groups do not differ from each other by double loci. DLVs were also found in singletons which don't form clonal complexes. The DLVs can be found between two singletons, between singleton and group and between groups. ST 503 and ST518 were singletons which differ from each other by DLV.ST 503 and ST518 belongs to *emm* 86.2. ST493, 494,522 which forms group 8 has DLVs with group 12 (which does not have bootstrap value) comprising ST515 and 495. Both these group belongs to *emm* 110. ST119 found in group6 has DLVs with singleton ST488 and also with ST532 which was found in group 19. ST119 and ST488 belongs to *emm* 100 while ST532 belongs to *emm* 11.

Sixty one sequence types were obtained out of seventy hospital GAS isolates **(Figure 12)**. With the standard definition of an eBURST group, all ST's assigned in the same group must share alleles at least six of the seven house- keeping loci with at least one other in ST in the group. This identifies clusters of closely related genotypes that are considered to have descended from the same founder and are defined as a clonal complex. Based on this definition, there were 11 groups which had predicted founders for three groups which had low bootstrap values and 36 singletons. The rest of the groups did not have predicted founders. Group 1 clonal complex comprised of ST109, ST484 and ST514 where ST 109 was the predicted founder with bootstrap value 33%.The ST in this group belonged to *emm* 85. ST109 had 2SLV, ST484 and 514 had one SLV and one DLV each.

Group 2 comprised of ST228, 517and 478. ST228 was the predicted founder. This clonal complex was found between STs which belong to *emm* 49.4 (ST228 and478) and stKNB6 (ST517) which is a new type.ST 228 has only 2SLVs while ST517 and 478 had one DLV

and one SLV. Group 3 clonal complex comprised of ST494, 493and 522 has multiple candidate for predicted founder.

The 73 GAS isolates from the Community had 46 sequence types **(Figure 13)** .Based on eBURST definition, there were six groups with 4 predicted founder ST 477,525,491 and 534.Singletons accounted for 30 numbers. ST477 was the predicted founder with bootstrap 31% for group 1 which had ST477, ST499 and ST500. ST477 had only 2SLVs while ST499 and 500 had one SLV and one DLV each. Groups 1 belongs to *emm* 44(ST477) and 113(ST499 and 500).

The second clonal group from the community comprised of ST525, 109 and 528.The predicted founder was ST525 with bootstrap value 32%. The third clonal group belongs to *emm* 105, the STs were ST491, 151 and 492 with predicted founder ST491.Group 4 had ST 534 as predicted founder, the group had ST534, 591 and 529 as clonal complex. This complex belongs to *emm* 49.4.Group 5 and 6 has no predicted founder; group 5 consisted of ST 535 and 526 while group 6 consisted ST 496 and 532.

More clonal groups were found in hospital eBURST rather than Community GAS isolates eBURST, but predicted founder were more from community rather than Hospital GAS isolates eBURST. With the eBurst data and the population snapshot, the sequence types from both community and the hospital GAS isolates which shared close relationship were inferred. There are clonal complexes found within GAS isolates from the community and the hospital GAS isolates but it is found more in community than hospital.

6.13. Phylogeny of GAS isolates based on their *emm* types

A phylogenetic tree was constructed based on 500 base pairs of 49 representative *emm* types which were selected for multi locus sequence typing. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 16.07488345 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4[226] **(Fig.14)** .

The tree identifies two main branches. The first branch was very small with 6 types while the second branch was very large and consisted of all remaining *emm* types. However, *emm* gene characterization alone does not provide any measure of genomic differentiation between GAS isolates because of high frequency of intragenic (point mutations) and intergenic recombination (deletion or addition of segments of chromosomes)

6.14. Phylogeny of GAS isolates based on their MLST profiles

A phylogenetic tree was constructed based on the 93 ST's of MLST. For this, a concatenated sequence was obtained by fusing all the seven loci's from the allelic profile **(Fig.15)**. The concatenated sequences formed were of 3,314 base pairs. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.17490760 is shown. The tree is drawn to scale, with branch lengths in the

same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding There were a total of 3134 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. *ska*

The results from the tree showed a very similar picture as that of the *emm* tree. There were two main branches, of which one was very small with only three ST's. The second and larger branch had all the remaining ST's. The latter consisted of many smaller clusters where the ST's were very closely packed.

When this ST types were compared to their respective *emm* types to ascertain the association between *emm* type and STs. Sequence types are very clonal within *emm* types except *emm*1 (ST36, 469, 30, 15), *emm* 12 (ST36, 28, 472), *emm* 18 (526,120), *emm* 55(100,528) . That is the clustering of ST's was in good agreement with *emm* types. The results when compared with eBURST results they showed similarity. The high degree of concordance between sequence types of MLST and *emm* type provides strong evidence that the MLST scheme gives accurate information or identifications of clones or clonal complex.

6.15. Linkage disequilibrium Of the GAS isolates from community and Hospital

One forty three GAS isolates representing 49 *emm* types and 93 STs were used to analyze whether the alleles at different loci within a population of GAS are randomly associated (linkage equilibrium), or whether there is a significant association between the alleles at different loci (linkage disequilibrium). The observed variance in the distribution of allelic mismatches in all pair-wise comparisons of the allelic profiles was computed and compared

to that expected in a freely recombining population. The extent of recombination within the GAS population was assessed by the index of association. The index of association can be calculated by the formula.

$$\mathbf{I_A = V_O / V_E - 1}$$

Ia = 0.367688
Observed variance (V_{obs}) 1.140165
Expected variance (V_{exp}) 0.833643
Mean trial variance = 0.985549
Max trials variance = 1.111028

Significant linkage disequilibrium is established if the observed variance, obtained with the real data set, is greater than the maximum variance observed with any of the randomized data sets with p values 0.05, 0.01 or 0.001. Index of association will not be significantly different from zero since $V_{O \text{ and }} V_E$ should be approximately the same in a population with linkage equilibrium. If the observed variance is less than that obtained with the randomized data sets there is no deviation from linkage equilibrium. The observed variance (V_{obs}) was greater than the maximum variance obtained in 100 trials**.** Significant linkage disequilibrium was detected. The index of association can also be calculated at www.mlst.net where online soft-wares are available.

7. DISCUSSION

Group A streptococcus (GAS) is a common human pathogen that is usually transmitted through the nasopharyngeal reservoir. Its clinical manifestations range from superficial infections, such as pharyngitis or impetigo, to invasive diseases, such as bacteremia and necrotizing fasciitis (NF), which may be accompanied by toxin-mediated manifestations, such as streptococcal toxic shock syndrome (STSS). A surge in the incidence and severity of GAS infections since the 1980s has been documented in many industrialized countries. These increased rates appear to have persisted well into the 1990s, and they have been largely attributed to the emergence of GAS strains with high potential for causing severe human infection. Certain studies have pointed to an association between particular M serotypes (or *emm* genotypes) and severe GAS infection. A number of studies have also suggested that serious GAS disease is associated with the emergence of highly virulent GAS clones whereas others have found little or no evidence of this association.

The population genetic structures of GAS infection from Southern India are very poorly reported. So our study was designed to explore the genetic structure of GAS causing human infections in Southern India using multi locus sequence typing. *emm* typing provides the basis for the typing scheme. M proteins are one of the key virulence factors in causing GAS disease. Antigenic differences in the hyper variable region constitute the basis for the Lancefield classification of GAS which has been expanded after the cloning of M-protein (*emm*) gene and the standardization of *emm* typing based on the sequencing of the amino terminal nucleotide residues. Multi locus sequence typing (MLST) is a nucleotide sequence method for characterization of the genetic relationship between the organisms of bacterial

species and uses housekeeping genes. Unlike housekeeping lo, *emm* genes are highly variable. Therefore *emm* type would change more rapidly than alleles at housekeeping loci resulting in variation within *emm* gene among isolates of clonal complex.

In our study, *emm* typing of 319 GAS isolates from the community identified 73 *emm* types on the whole, which were from four different sources namely pharyngitis, normal throat, impetigo and normal skin. Forty five *emm* types from community pharyngitis, forty seven *emm* types from Normal throat, thirty five *emm* types from impetigo and twenty five *emm* types from skin colonizers were identified. This confirmed the high heterogeneity among them. There were 19 sequence types and two new types stKNB3 & stKNB8 were identified among the community GAS isolates.

The most common *emm* types were restricted to few types and differed from the site of isolation namely pharyngitis, normal throat, impetigo and normal skin. From the cases of pharyngitis *emm* 11 was the dominant while *emm*112 was the most common among normal throat and impetigo. In normal skin, *emm* 93 was the most common *emm* type. There were few types seen predominantly associated with pharyngitis, one such *emm* type is *emm* 110 which accounted for 8 of the GAS isolates. *emm* 112 the most common among the seventy three *emm* types which accounted for 29 (9 %) of the GAS isolates and was found predominantly in normal throat *(n=18)* and impetigo skin *(n= 8).* Some impetigo types were also seen as colonizers in throat of asymptomatic children. They were *emm* 11, 28, 44, 49, 55, 63, 82, 85, 92, 100, 105,111, 112, 113,119 and st854. So normal throat acts as a GAS reservoir in the community and harbors strain capable of causing disease. The colonization may occur in individual with GAS *emm* types and manifests as infection or carriage; it all depends on the individual's state of immunity, genetic constitution and the environment **.**

Studies from North India indicated out of the 94 GAS isolates when *emm* typed, thirty seven *emm* types were identified. The most frequently observed *emm* types were *emm*49 (8.5%) and *emm*112 (7.5%) followed by 6.5% of *emm* 1-2, *emm* 75 and *emm* 81. The predominant *emm* types from throat (*emm*49 and *emm*75) were different from those of skin (*emm*44, *emm*81 and *emm*112) samples. Despite the diversity in the *emm* type pattern of throat and skin isolates no significant association between *emm* type and source of isolation was observed [20, 268] .A study from Chennai revealed 22 *emm* types out of 34 GAS isolates, in which the most common type was found to be *emm* 49 followed by *emm* 82 and *emm* 74 [269]**.**

The most common *emm* types among the South Indian GAS isolates were not the dominant types from North India though they had some similarities. *emm*49 and *emm*75, the most dominant *emm* types from throat GAS isolates (North India) were found in the throat GAS isolates from South India but increasing numbers were found in the throat colonizers only, while *emm* 11 was the most common *emm* type in our study. *emm* 112 the dominant *emm* type from our studies were found in the North Indian studies and it was found to be the most common *emm* type from the Skin isolates from our study and from the reported literature from the North India. The *emm* types reported in Chennai were also identified in our studies.

The most frequently observed *emm* types from our study were not only different from previously reported most prevalent *emm* types in India but also from epidemiological studies of isolates from other countries like Japan , Taiwan , Germany , Australia and United States . The difference in the most prevalent *emm* types from this study in

comparison to earlier Indian reports could be due to the fact that the most prevalent *emm* type within population change over time, which can be predicted by continuing surveillance. From the foreign literatures there were reports indicating the association of particular *emm* types that have been strongly associated with GAS invasive disease and some *emm* types with acute pharyngitis and impetigo. The individual strains and clones that cause infections and their association between their genotypes and *emm* types have been studied in numerous laboratories **.**These studies showed that invasive disease in United States and developed countries is caused by small number of GAS clones. In United States approximately 30% of cases of invasive GAS diseases that occurred during the 1990s were caused by isolates of M1 and M3 serotypes.

From our study, we found that there was no association between the *emm* types and the site of isolation from the community and the GAS isolates from the hospital, but the *emm* types from the community pharyngitis were found more among invasive GAS isolates from hospital rather than the *emm* types found in the hospital pharyngitis. The community pharyngitis *emm* types found in the Invasive GAS isolates were *emm* 58, 63, 74, 85, 86, 93, 102, 105 and 110.So there is only minimal or no evidence to support the contention regarding the highly virulent GAS clones from Southern India. This is also supported by the study from North India where there was no association found between *emm* type and the source of isolation.

Understanding the extent of importance of genetic and biochemically diversity among strains of the same or very close related species is a corner stone issue in microbiology disciplines, such as taxonomy, diagnosis, epidemiological studies diversity surveys, biogeography etc. The multi locus sequence typing method has emerged as method of

choice for exploring and cataloguing intra species genetic diversity, thus setting the stage for linking the genetic diversity to the biochemical and functional diversities of species. Typical applications of MLST method employ the sequencing of seven genes and subsequent phylogenetic analysis concatenated sequences, alignments to reveal the exact genetic relationships among the strains analyzed.

In India; this is the first study using the technique multi locus sequence typing. From our MLST studies we confirmed that heterogeneity among GAS isolates is associated with allelic variations in seven housekeeping genes. Out of the 143 GAS isolates subjected to MLST, 93 sequence types (ST) were identified from 49 *emm* types**.** Among this 93 sequence types, 68 were new sequence types and 25 were known sequence types. High percentage of STs was hitherto unrecognized types. This shows that large number new genotypes develop in this community, probably because of highly susceptible hosts in the population. This new type also indicates there is constant evolution within the GAS isolates, though mutations within housekeeping genes are assumed to be slow or neutral.

In light of the initial published data of MLST for GAS, where 212 GAS [38] isolates with 78 *emm* types revealed hundred sequence types (1-100STs) and their isolation of human subjects dated back from 60 years. With a time period spanning 60years and 78 *emm* types, hundred sequence types were identified. From our study we were able to identify 93 sequence types out of 49 *emm* types from 143 GAS isolates collected in a short span of time.

In comparison to the initial study of GAS MLST of Spratt and Debra Bessen [38], there were twenty three *emm* types which were also found in our study out of which only two *emm* types had similar sequence types (*emm* 12-ST36 and *emm* 93-ST10) and rest of the

emm types had different STs. According to another published data, where 137 GAS isolates from remote community of aboriginal community from Australia [252] representing 31 *emm* types subjected to MLST identified thirty three sequence types. In this study also revealed in comparison with our data showed that there were only few *emm (n=4)* types which shared similar STs, they are *emm* 100(ST119), *emm* 44/61(ST178), *emm* 85 (ST109) and *emm* 60(ST193).

Published reports from Nepal [125] in relation to MLST of GAS revealed 45 *emm* types out of 120 noninvasive GAS isolates with fifty one STs. There were twenty five *emm* types found in Nepal study in relation to our MLST study, out of which only few *emm* (*n=12*)types shared same STS. They were *emm* 25.2 (ST350), *emm* 44 (ST351), *emm* 49.4 (ST228), *emm* 55 (ST100), *emm*66 (ST249), *emm*77 (ST347), *emm*82 (ST320), *emm*85 (ST109), *emm*100 (ST119), *emm*104 (ST353), *emm*105 (ST151) and *emm*119.2 (ST237). In general, sequence type from Nepal and our studies indicate that there are more STS common than other reported literatures from Australia [252] , United States [38], Germany [254] , London[253],and Portugal ; this could be due to geographical proximity. This clearly describes that the allelic profiles from our studies are very novel though they share *emm* types and housekeeping alleles with isolates from worldwide sources.

More frequently, certain *emm* types showed multiple STs, showing how allelic variations among house keeping genes can alter the genotype of the same *emm* type. Out of forty nine *emm* types there were twenty five *emm* types with more than one sequence type .They are *emm* 1, 12, 18, 28, 44, 49.4, 55, 74, 75, 77, 85, 86, 92, 93, 100, 105, 108, 110, 111, 112, 113, 118, st6735, st212 and st854.1. *emm* type 110 and 49.4 had five STs identified among them while six *emm* types 1, 18, 44, 105, 100 and 85 had four different sequence types, Four *emm*

types namely 92,113, 108 and 12 had three different STs and thirteen *emm* types had two different sequence types. Sixty five of the 93 sequence types were represented by single isolate. Genotypic variations had no association with the site from where the GAS isolates were recovered. This shows that these variations are not site-specific. Probably they undergo such variations to overcome any host related factors during their initial stages of colonization.

The different sequence types within same *emm* type from our study when checked for maximum nucleotide identity; they had 99% nucleotide identity except *emm* 108 which had 98% identity within STs. The allelic profiles of different sequence types within an *emm* type were analysed .Most of these allelic variations seen among these sequence types were single locus variations. Few were also double locus variations. But multi locus variations were rare. This shows that the allelic variations are due to random point mutations in the house keeping gene.

Certain sequence types were found to have more than one *emm* type. They are known as *emm* variable sequence types. They were ST 36, ST 100, ST 357 and ST 477. *emm* type 1 and 12 shared ST 36, *emm* type 55 and 92 shared ST 100, while *emm* type 75 and 108 had ST 357 and *emm* type 44 and 113 shared ST 477. Pair wise comparison of *emm* types sharing the same STs were done using clustal W .The maximum nucleotide identity was identified as follows ;(ST 36) *emm* type 1 and 12 had a percentage identity of 50.47%, (ST100) *emm*55 and 92 had percentage identity of 53.42%, (ST 357) had % identity of 56.73% and (ST477) *emm* 44and 113 had percentage identity of 62.96% . The different *emm* types having same STs were mostly isolated from hospital pharyngitis and impetigo from community. From the reported literature of Awdesh Kalia, Karen, Debra Bessen and Spratt

on MLST for 495 GAS isolate [38] representing 158 *emm* types with two hundred and thirty eight sequence types, there were twelve *emm* variable sequence types. They were ST65 (*emm*19,29and 30), ST83(*emm*17 and 47), ST84(*emm*14 and 51), ST3 (*emm*33 and st211), ST4(*emm*43 and52) ,ST9(*emm*86 and 97), ST10 (*emm*70, 80, 93, 98, 108 and 121), ST11(*emm*53 and101),ST123(*emm*115 and 123),ST174 (st2911and st3850), ST182(*emm*101 and st5282) and ST39(*emm*4 and st104). The extent of similarity between *emm* sequences of those isolates that have the same ST but different *emm* types distinguishes variation in *emm* type. The large sequence difference between *emm* types strongly suggests horizontal gene transfer (HGT) followed by intergenomic recombination rather than point mutation [270] **.** The *emm* variable STs seen in Southern India was not found in the above reported literature which also supports that the population genetic structure of GAS is different for different parts of the world.

In multi locus sequence typing, the relatedness among isolates is displayed as dendrogram, based on differences in allelic profiles, which identifies clusters of similar genotypes. Groups of isolates with closely related allelic profiles are known as clonal complex. Within each clonal complex it is usually possible to identify one predominant allelic profile and a number of less common variant allelic profiles. The simplest explanation for this pattern is that the predominant allelic profile (predicted founder) represents the genotype of the ancestral clone that give rise to the clonal complex, and the variants differ from this allelic profile at a single locus (SLV-single locus variants) represents the initial stages of diversification of clone. The ancestral clone can be defined using the BURST program which indicates the allelic profile within each clonal complex that differs from the largest number of other allelic profiles at only a single locus and, hence, is most likely to be

phylogenetically central [41] .In eBURST a clonal complex is defined as a group of STs in a population that share 6/7 alleles with at least one other ST in the group. eBURST identifies clonal complex and displays the evolution of STs within clonal complex from the predicted founder or founder ST The predicted founder ST has increased number of Single locus variants (SLV) supported by bootstrap values.

The population snapshot of eBURST clearly represents the clonal groups found among the 93 sequence types of 143 GAS isolates. There were 19 groups with eight groups having the predicted founder out of which two had bootstrap value more than 60%.This 19 groups infers that two third of the allelic profile or the sequence types are clonal in nature Among the 19 groups, there are also groups without predicted founder which does not explain evolutionary descents. The group 1 clonal complex with predicted founder ST534 having a bootstrap value of 70% comprising of ST534, ST228, ST517, ST531, ST529 and ST478 belonged to *emm* type 49.4 except ST 517 which belonged to newly identified *emm* type stKNB6.ST517 is a single locus variant of ST534 which represents the relationship between *emm* 49.4 and stKNB6 indicating they are clonal in nature.

The second group of clonal complex of the eBURST **e**xplains the clonal complex of five sequence types ST109, ST525, ST484, ST514 and ST 528 which belong to *emm* 85 except ST528 which belong to *emm* 55. ST 109 being the predicted founder differs from the ST 528 by DLV. eBurst clearly represents the clonality between two different *emm* types 85 and 55.The approach used in eBURST simplifies the problem of depicting the evolutionary relationship among closely related genotypes that are poorly represented on a tree. eBURST also depicted the relationship or clonality of sequence types found within each group and between different *emm* types within the same group namely group 4(between *emm* types 74

and 18), group 5(between *emm* 44 and113) group 6 (*emm* 100), group 7(*emm*18),and group8(*emm*110). eBURST does not make inference about the relationship about the distantly related genotypes that belong to different group due to the extent of homologous recombination which is high [263].There was one *emm* variable ST(ST477) which formed a clonal complex identified by eBURST . Single locus variants (SLV) are found only in the clonal groups, whereas double locus variants (DLV) are found between the clonal groups, between singletons and between singletons and groups. There were 35 pairs of SLV among 143 GAS isolates from hospital and community within the 19 clonal groups. From the eBURST it can be clearly understood that two third of the GAS isolates are clonal in nature from Southern India which causes infection.

According to the BURST algorithm, a genotype occasionally increases in frequency, as a consequence of genetic drift or natural selection, and diversifies by the accumulation of mutations or localized recombinational replacements, to result in slight variants of the founding genotype. In the context of MLST, the members of this emerging clone will initially be indistinguishable in allelic profile (same ST) but overtime they will diversify to produce a number of variants in which one of the seven MLST loci has been altered and forms SLVs. Further diversification will produce variants of the founding ST that differ at two of the seven loci and forms DLVs. Bacterial population should therefore consist of clonal complexes which provided they are new(young) complexes, should easily be apparent using multi locus approaches to strain characterization. According to MLST, the founding ST identified should contain increased number of SLVs. The predicted founders among 143 GAS isolates are ST534, ST109, ST120, ST477, ST119 and ST535. ST534 had 4SLVs with bootstrap value 70% and average genetic distance of 1.2.Bootstrapping provides degree of

confidence in the assignment founder of clonal complex. Another useful measure provided by BURST is the average genetic distance (measured as the average number of allelic differences) of each ST from all other ST's in a clonal complex. The ST within a clonal complex that has minimum average distance is likely predicted founder [263]. ST 109 had three SLV with average genetic distance 1.25 which was less than other sequence type's average genetic distance within clonal complex. ST120, ST477, ST119 and ST535 had two SLVs with average genetic one within their respective clonal complex. So the predicted founders found in BURST for 143 GAS isolates are the likely predicted founders.

BURST algorithm was able to identify the young clonal complexes among the ninety three sequence types from 143 GAS isolates. The simple or young clonal complex has a predicted founder with SLVs; the average genetic distance from the founding ST to all other STs in the clonal complex will be one, and all other ST will have a greater average distance [263]. There were three clonal complexes which matched the criteria for young complexes; they were ST120, ST477 and ST535.This sequence types had an average distance of one, and all other sequence type had 1.5 as their average distance.

When the clonal complex with ancestral clones and their associated SLVs have been identified, it is possible to distinguish SLVs that have arisen by recombination from those that have arisen by point mutation. These distinctions are made on the basis of two criteria: the number of nucleotide sites that differ between the variant allele in the SLV and the typical allele in the putative ancestral clone, and the frequency at which the variant allele is found elsewhere in the database. The occurrence of multiple point mutations at one locus with no changes at the other six loci is unlikely, and variant alleles that differ at multiple nucleotide sites are therefore considered to be recombinational replacements [271].

However, differences at a single site could result from either point mutation or from recombinational replacements between very similar sequences that introduce only a single nucleotide difference. These types of events can be distinguished, as a point mutation within a house keeping locus is very likely to result in a variant allele that is unique within MLST database.

There were two clonal complexes among the 143 GAS isolates with ninety three sequence types showing bootstrap value more than 60% ,with predicted founders ST534 and ST109 from BURST analysis, were checked for the number of nucleotide site difference between the variant allele within complex. Group1 clonal complex comprised of sequence types ST534, ST 228, ST478, ST517, ST529 and ST531. ST534 was the predicted founder. ST534 was found to be SLV of ST228, ST517, ST529 and ST531 and double locus variant for ST478. ST228 and ST517 differed from ST534 at the loci *muts* (DNA mismatch repair protein). ST529 and ST 531 had a variable allele at the loci *gtr* (glutamine transporter protein) from ST534. ST478 a DLV, differed from ST 534 at the locus *mutS* and *gtr*. ST534 had an alignment score of 99 with ST 228 for the loci *muts* when checked for pair wise alignment.ST534 had the allele number 3 while ST228 had allele number 11 for the loci *mutS.* These two alleles when compared for nucleotide difference, they differed from each other by three nucleotides. Similarly the allele number of ST 534 and ST517 at the loci *mutS* showed a nucleotide difference of two and their respective allele numbers were 3 and 5. ST534 when compared with ST529 variable allele at the loci *gtr*, had three nucleotide difference.ST534 and ST531 had varied allele which differed by two nucleotide, the allele number were 31 and 35 respectively for the loci *gtr*. All the SLV of ST534 differed from each other by nucleotide difference of 2-3 nucleotides and had alignment score of 99.

Group 2 clonal complexes comprised of sequence type ST109, ST484, ST514, ST525 and ST528. ST109 was the predicted founder. ST109 had three SLVs; they are ST484, ST514 and ST 525. ST528 was a DLV of ST109. ST109 varied at the loci *gki* (glucose kinase), *gtr* (glutamine transporter protein), *mutS* (DNA mismatch repair protein) and y*qiL* (acetyl CoA acetyl transferase) with rest of the sequence types. ST525 differed from ST109 at the loci *gki*. ST484 differed from ST109 at the loci *gtr*. ST528 differed from ST109 at the loci *mutS*. ST514 differed from ST 109 at the loci y*qiL*. ST528 differed from ST109 at the locus *gki* and *mutS*. These SLVs of ST109 were checked for nucleotide difference with the respective locus. The ST 109 with ST 525 for the locus *gki* had a nucleotide difference of one and it was the same between ST109 and ST528. ST109 with ST484 had a nucleotide difference of one with loci *gtr*. With loci y*qiL*, ST 109 and ST514 the alleles varied by a nucleotide difference of two. ST109 and ST 528 which is a DLV had a maximum nucleotide difference of five.

The nucleotide differences in association with the predicted founder within the clonal complex which had bootstrap value more than 60% showed that the varied alleles or alleles which differed from the predicted founder at different locus with other sequence types differed by a nucleotide difference of one to five. This clearly shows that the allelic variation can be due to point mutation which has been accumulated over time or due to recombination. Although we cannot, at present estimate the rate of recombination in nature, when combined with strong selective pressure, even relatively low rates of recombination can play a fundamental role in accelerating the evolution of functional diversity.

The clonal complexes present in the community and the clinical GAS isolates were revealed using BURST algorithm. There were 46 sequence types identified within seventy three

community GAS isolate. Using the stringent definition of eBURST the population snap shot of community GAS isolates was obtained. There were six groups out of which only four clonal complexes had the predicted founder. The predicted founders are ST525, ST477, ST491 and ST534.The average genetic distance of the predicted founder sequence type was checked in respect to the other sequence types found within the clonal complex, it was found that three clonal complex were young clonal complex. The predicted founders which had young clonal complex were ST525, ST477 and ST491. Community GAS isolates had thirty singletons among forty six sequence types. Whereas the clinical GAS isolates population snapshot of BURST with seventy GAS isolates and sixty one sequence types showed that there were eleven groups with two predicted founder among them. The two predicted founder were ST109 and ST228. The clonal complex formed by the predicted founder ST109 and ST 228 were young clonal complex and thirty six singletons were found.

In relation to clinical the community had forty six sequence types and four clonal complexes with predicted founder, whereas there were only two clonal complexes with predicted founder out of sixty one sequence types in the clinical. Community GAS isolates had only six groups in comparison to clinical GAS isolates which had eleven groups. There were two young clonal complexes in the clinical GAS isolates and four young clonal complexes in the community GAS isolates. These findings may indicate that the evolution occurs more in clinical GAS isolates than community GAS isolates. The reasons being more new sequence types were observed in clinical GAS isolates: Secondly more clonal groups were found in clinical GAS isolates (with or without predicted founder); more SLVs were also found in clinical isolates.

Molecular phylogenetics have become an integral part of research endeavors in diverse areas of molecular biology, population genetics, developmental biology and evolutionary biology and has implications for ecology and medicine. The observed heterogeneity of evolutionary rates among lineages in a gene is caused partly by the nondeterministic nature of the evolutionary process, partly by differences in intensity and type of natural selection and partly by unknown factors. Tree building methods, with the exception of UPGMA, don't assume constancy of evolutionary rates vary significantly among lineages. However some methods are known to produce consistently incorrect results when the evolutionary rates vary significantly among lineages. Felsenstein showed that if a four sequence tree contains two long and two short branches, then the long branches tend to cluster together in the maximum parsimony (MP) trees even if they are distantly related (long-branch attraction problem). One to avoid this problem is by using a large number of sequences such that the long branches are broken. Another way to minimize the effects of long-branch attraction is to use maximum likelihood method (ML).

The phylogenetic analyses based on the concatenated sequences of ninety three sequence types of MLST were conducted using the molecular evolutionary genetics analysis software (MEGA 4) [271].The neighbor-joining (NJ) method was used which is widely used in reconstructing phylogenies because of its computational speed and the high accuracy in phylogenetic inference. The tree size does not influence the efficiency of NJ in constructing shallow and deep branches, in which the evolutionary process is assumed to be homogeneous in all lineages. Maximum likelihood method is generally regarded as the gold standard of molecular phylogenetic reconstruction by using which the evolutionary distance[272].

The ninety three sequence types were used for construction of the phylogenetic tree. Several clusters were formed among genotypes which were closely related. The STs within the clusters were compared to their respective *emm* types to ascertain the association between *emm* type and ST. The STs were found to be clonal in respect to their *emm* types. The cluster containing ST534, ST228, ST517, ST529, ST531 and ST478 all fall under a single *emm* type 49.4 except ST517 which belongs to stKNB6. Many clusters were found in which the STs was in good agreement with *emm* types, but there were some exceptions. They are *emm*1 (ST36, 469, 30, 15), *emm* 12 (ST36, 28, 472), *emm* 18 (526,120) and *emm* 55(100,528). In tree the evolution of STs was not very clear as in BURST snapshot. The approach used in BURST greatly simplifies the problem of depicting the evolutionary relationship among closely related genotypes, which are poorly represented on tree. So eBURST displays the relatedness among GAS isolates very clearly.

The population structures of bacterial species are complex and often controversial. The existence of clones within bacterial populations, and of linkage disequilibrium between alleles at different loci, is often cited as evidence for low rates of recombination.

The linkage disequilibrium was calculated from the distribution of allelic mismatches between all the GAS isolates using the index of association. Significant linkage disequilibrium was detected because the observed variance was found to be greater than the expected variance which indicates there is association between allele at different loci in the seven housekeeping genes.

8. SUMMARY AND CONCLUSION

1. Seventy three *emm* types were identified among the 319 GAS isolates from the four different sites from community, thirty *emm* types from 41 GAS isolates of hospital pharyngitis and 37 *emm* types from 58 invasive GAS isolates, this indicates the role of multiple *emm* types in causing human infection. This also clearly explains the absence of association between *emm* types and their site of isolation and depicts the high heterogeneity seen among the GAS isolates from South India.
2. Some *emm* types which cause impetigo were also seen as colonizers in throat of asymptomatic children. So normal throat acts as a GAS reservoir in the community and harbors strain capable of causing disease.
3. The dominant *emm* types found in South Indian community were different from the most common *emm* types reported from other regions of the world.
4. The difference in the most prevalent *emm* types from this study in comparison to earlier Indian reports could be due to the fact that the most prevalent *emm* type within population change over time, which can be predicted by continuing surveillance.
5. Despite the diversity in the *emm* type pattern of throat and skin isolates no significant association between *emm* type and source of isolation was observed.
6. Multi locus sequence typing of 143 GAS isolates identified ninety three sequence types out of which sixty eight were new sequence types. ST338/*emm* 63 was the dominant type followed by ST00/*emm*55&92.
7. Multi locus sequence typing studies revealed new sequence types which had the same alleles but arranged in different combinations. These different combinations in

turn formed sixty eight novel sequence types which were not found in other parts of the world.

8. The multi locus sequence typing using the house keeping genes showed that this technique can identify heterogeneity among individual *emm* types.
9. The sequence types within an *emm* type found in our study were also found in Nepal in higher numbers compared to the studies from Australia, United States of America, Germany, Portugal and London. These finding highlights that the population genetic structure of GAS isolates varies based on geographical proximity.
10. Genotypic variations had no association with the site from where the GAS isolates were recovered. This shows that these variations are not site-specific. Probably they undergo such variations to overcome any host related factors during their initial stages of colonization.
11. High percentage of sequence types from MLST were new (68 / 93STs) types. This shows that large number new genotypes develop in this community, probably because of highly susceptible hosts in the population.
12. The eBURST analysis showed that about two thirds of 143 isolates subjected to MLST had varying degrees of clonality. Comparison of nucleotide sequences using clustalW showed 99% identity. Most of these allelic variations are single locus variations. Few were also double locus variations. But multi locus variations were rare. This shows that the allelic variations were more due to random point mutations than recombination in the house keeping gene loci.

13. eBURST analysis clearly indicates that two third of the GAS isolates indicated that they are clonal in nature.
14. Significant linkage disequilibrium was detected which indicates the presence of low level of recombination.

CONCLUSION

This study was looked at to study the clonal diversity of GAS causing human infections in Southern India (or) population genetic structure of GAS disease in South India. These facets of this study have been poorly reported from India. This is the first study in India utilizing the technique multi locus sequence typing (MLST) for GAS which increased the information on diversity of GAS infections. From our study we have found out GAS isolates in relation to *emm* typing, shows high heterogeneity. Our MLST studies indicated that, there are new sequence types found which were not reported in any part of the world. It also depicted the high heterogeneity found within individual *emm* types in respect to sequence types. There was no association found between sequence types and to their site of isolation. The sequence types found based on their allelic profile of GAS isolates clearly explained there is no association between their STs and their site of isolation. eBURST population snap shot identified the clonal complexes present in the GAS isolates from the community and the hospital isolates and indicated that two third of the GAS population are clonal in nature with young clonal complexes.

9. RECOMMENDATIONS

- What are the environmental factors that convert a non-invasive to invasive?
- What host factors are involved in the development of the susceptible to severe invasive disease?
- How does one integrate Horizontal Gene Transfer (HGT) in context to genetic exchange and to arrive at meaningful microbiological concepts?
- How does one reach the robust conclusion on the presence or absence of HGT?
- Role of Linkage disequilibrium- Index of association.
- Group A Streptococcus Vaccine for highly diverse GAS population.
- Exploitation of functional genomics and experimental models to develop GAS vaccine.

Fig.1. Distribution of *emm* types from Community Pharyngitis GAS isolates (*n=101*)

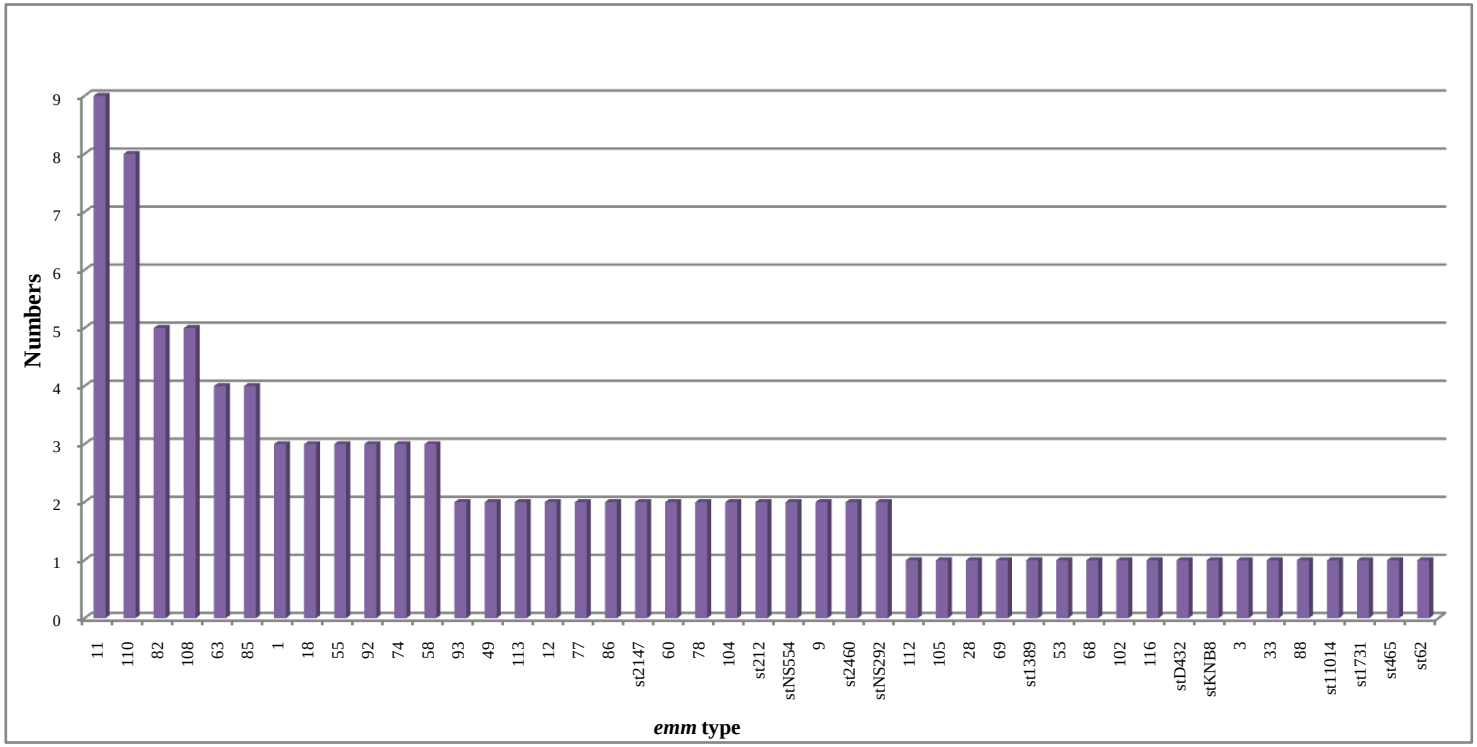


Fig.2.Distribution of *emm* types from community asymptomatic throat carriers GAS isolates (*n=113*)

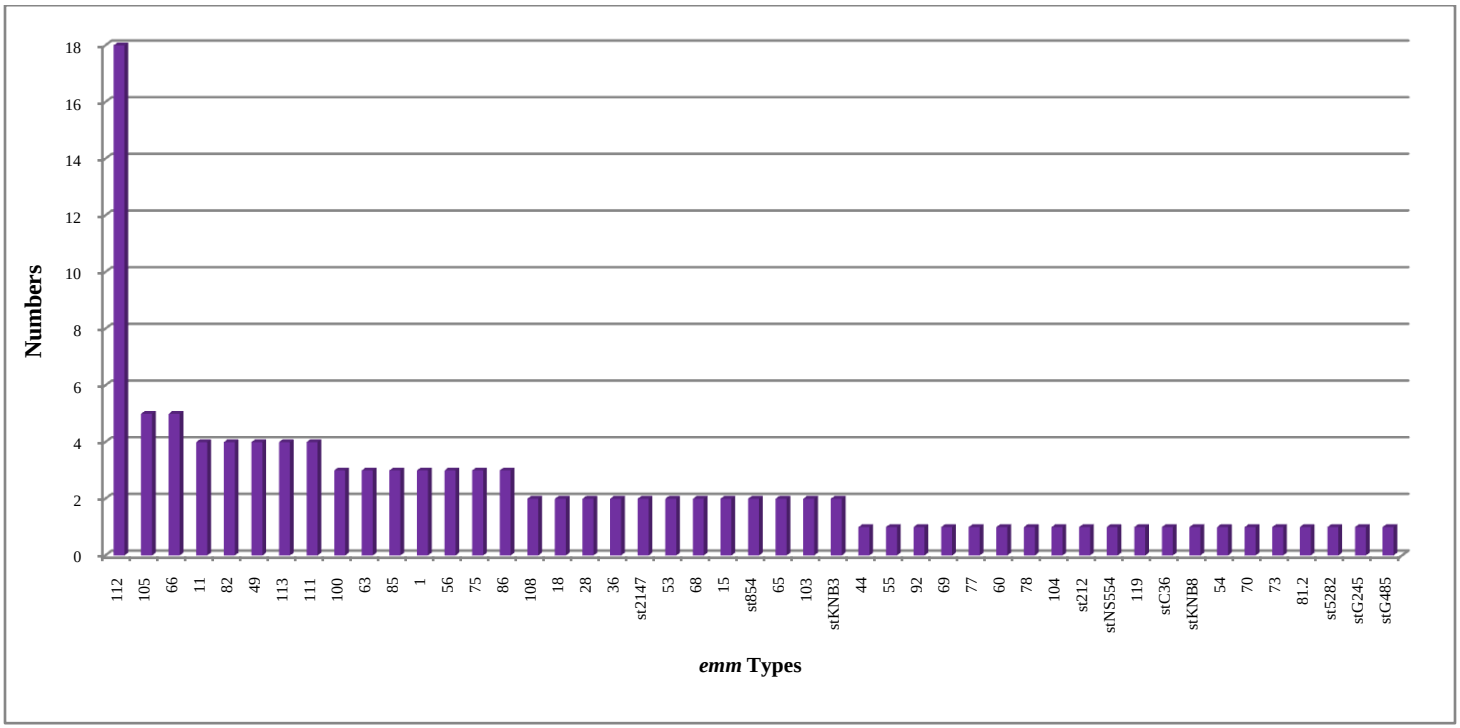


Fig.3.Distribution of *emm* types from community impetigo GAS isolates (*n=65*)

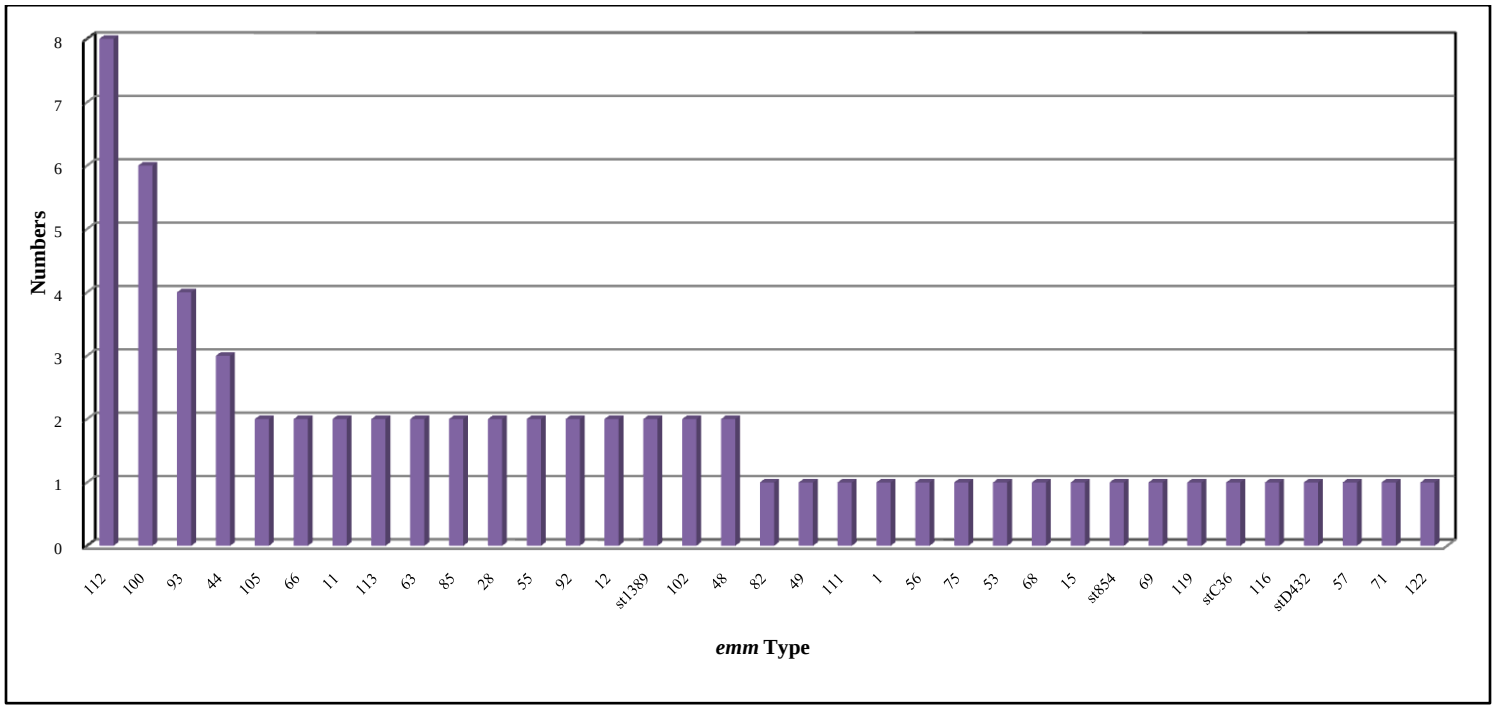


Fig.4.Distribution of *emm* types from community normal skin GAS isolates (*n=40*)

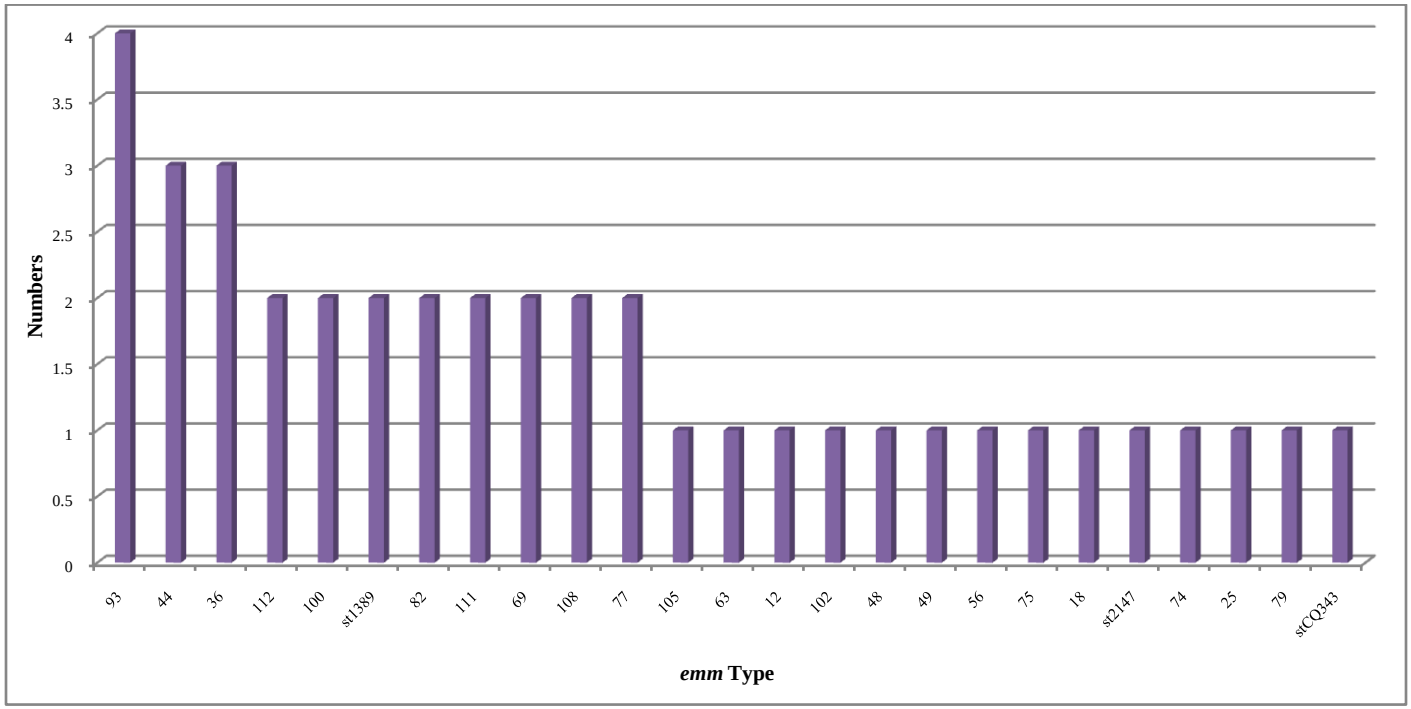


Fig.5. Distribution of *emm* types from Hospital pharyngitis GAS isolates (*n=41*)

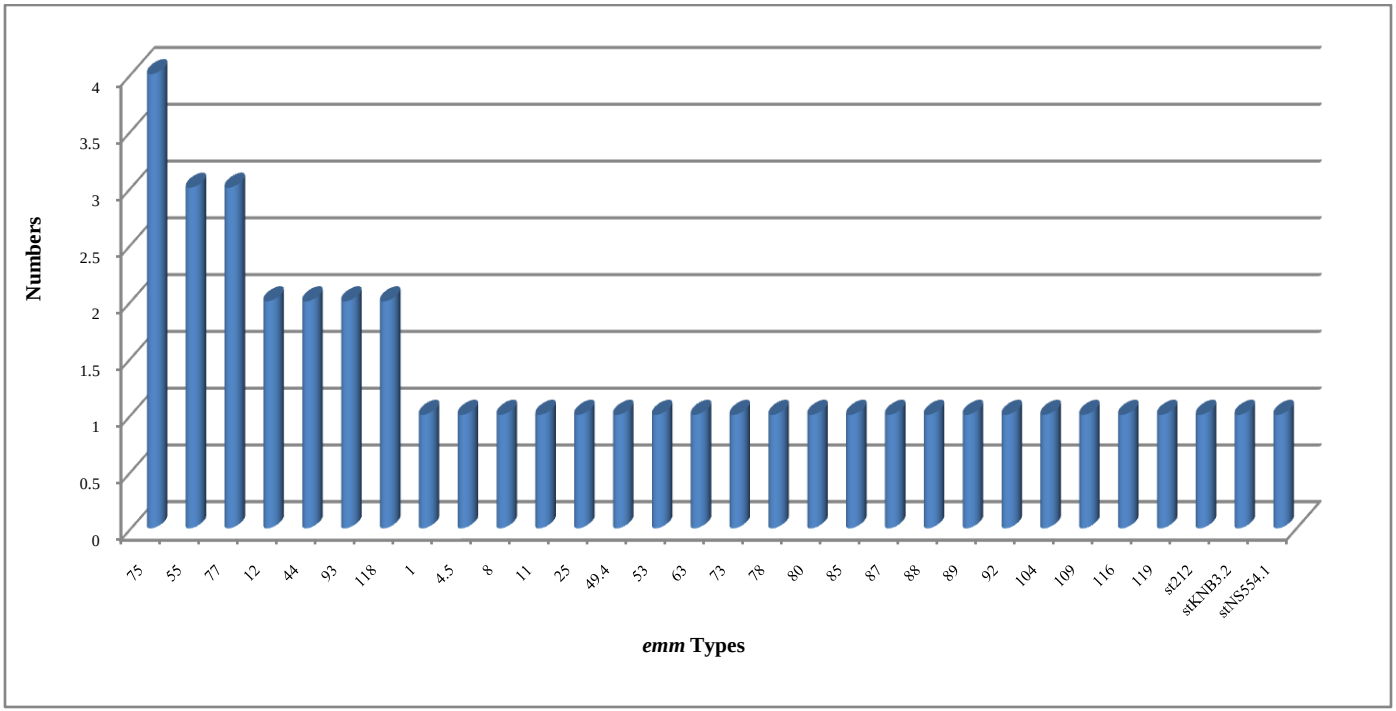


Fig.6. Distribution of *emm* types from invasive GAS isolates from Hospital (*n=58*)

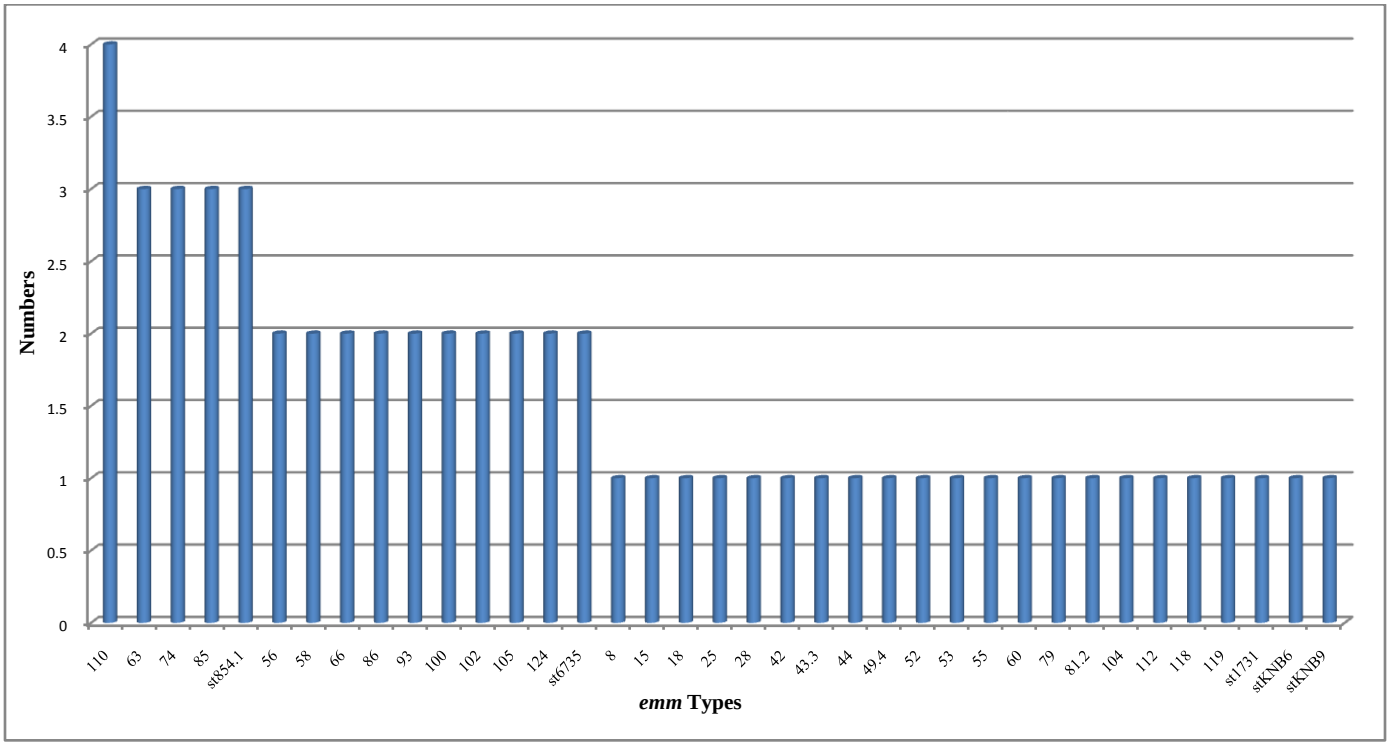


Fig.7.Multiple Sequence Alignment of ST 15, 36,469,530 of *emm* 1

```
ST469           AATCTTAATTGGAAAGAAACCCAAGAAGTCGGTTCCGTTGTTGAAAAAGAATTGGGCATT 60
ST530           AACCTTAATTGGAAAGAAACCCAAGAAGTCGGTTCCGTTGTTGAAAAAGAATTGGGCATT 60
ST36            AACCTTAATTGGAAAGAAACCCAAGAAGTCGGTTCCGTTGTTGAAAAAGAATTGGGCATT 60
ST15            AACCTTAATTGGAAAGAAACCCAAGAAGTCGGTTCCGTTGTTGAAAAAGAATTGGGCATT 60
                ** ************************************************************

ST469           CCTTTTGCCATTGACAATGATGCCAATGTGGCTGCCCTTGGTGAACGTTGGGTAGGTGCT 120
ST530           CCTTTTGCCATTGACAATGATGCCAATGTGGCTGCCCTTGGTGAACGTTGGGTAGGTGCT 120
ST36            CCTTTTGCCATTGACAATGATGCTAATGTGGCTGCCCTTGGTGAACGTTGGGTAGGTGCT 120
ST15            CCTTTTGCCATTGACAATGATGCCAATGTGGCTGCCCTTGGTGAACGTTGGGTAGGTGCT 120
                ********************** ************************************

ST469           GGTGAAAATAACCCAGATGTCGTCTTCATGACACTTGGAACAGGTGTCGGTGGAGGCATT 180
ST530           GGTGAAAATAACCCAGATGTCGTCTTCATGACACTTGGAACAGGTGTCGGTGGAGGCATT 180
ST36            GGTGAAAATAACCCAGATGTCGTCTTCATGACACTTGGAACAGGTGTCGGTGGAGGCATT 180
ST15            GGTGAAAATAACCCAGATGTCGTCTTCATGACACTTGGAACAGGTGTTGGTGGAGGCATT 180
                ********************************************** ************

ST469           ATTGCTGATGGTAACTTGATTCATGGTGTTGCAGGAGCAGGTGGTGAAATCGGCCACATT 240
ST530           ATTGCTGATGGTAACTTGATTCATGGTGTTGCAGGAGCAGGTGGTGAAATCGGCCACATG 240
ST36            ATTGCTGATGGTAACTTGATTCATGGTGTTGCAGGAGCAGGTGGTGAAATCGGCCACATG 240
ST15            ATTGCTGATGGTAACTTGATTCATGGTGTTGCAGGAGCAGGTGGTGAAATCGGCCACATG 240
                ************************************************************

ST469           ATTGTTGAGCCAGAAAATGGCTTTGCTTGTACTTGTGGCTCACACGGCTGTTTGGAAACA 300
ST530           ATTGTTGAGCCAGAAAATGGCTTTGCTTGTACTTGTGGCTCACACGGCTGTTTGGAAACA 300
ST36            ATTGTTGAGCCAGAAAATGGCTTTGCTTGTACTTGTGGCTCACACGGCTGTTTGGAAACA 300
ST15            ATTGTTGAGCCAGAAAATGGCTTTGCTTGTACTTGTGGCTCACACGGCTGTTTGGAAACA 300
                ************************************************************

ST469           GTAGCTTCAGCAACAGGAGTTGTCAAAGTGGCACGTTTACTGGCAGAAGCCTACGAAGGG 360
ST530           GTAGCTTCAGCAACAGGAGTTGTCAAAGTGGCACGTTTACTGGCAGAAGCCTACGAAGGG 360
ST36            GTAGCTTCAGCAACAGGAGTTGTCAAAGTGGCACGTTTATTGGCAGAAGCTTATGAAGGG 360
ST15            GTAGCTTCAGCAACAGGAGTTGTCAAAGTGGCACGTTTACTGGCAGAAGCTTATGAAGGG 360
                ************************************* ********** ** ******

ST469           GATTCAGCCATCAAAGCAGCTATTGACAATGGTGAAGGTGTTACCAGTAAAGACATTTTC 420
ST530           GATTCAGCCATCAAAGCAGCTATTGACAATGGTGAAGGTGTTACCAGTAAAGACATTTTC 420
ST36            GATTCAGCCATCAAAGCAGCTATTGACAATGGTGAAGGTGTTACCAGCAAAGATATTTTT 420
ST15            GATTCAGCCATCAAAGCAGCTATTGACAATGGTGAAGGTGTTACCAGTAAAGACATTTTT 420
                ********************************************** ***** *****

ST469           ATGGCGGCTGAAGCAGGGGATTCCTTTGCTGATTCTGTTGTGGAAAAGGTTGGTTACTAC 480
ST530           ATGGCGGCTGAAGCAGGGGATTCCTTTGCTGATTCTGTTGTGGAAAAGGTTGGTTACTAC 480
ST36            ATGGCAGCTGAAGCAGGGGATTCCTTTGCTGATTCTGTTGTGGAAAAGGTTGGTTACTAC 480
ST15            ATGGCAGCAGAAGCAGGGGATTCCTTTGCTGATTCTGTTGTGGAAAAGGTTGGTTACTAC 480
                ***** ** ************************************************************
```

```
ST469           CTTGGCCTTGCTTCAGCAACCTTACTTCGAACTATGAATCTTTTGGAAGTACCAACTAAG 540
ST530           CTTGGCCTTGCTTCAGCAACCTTACTTCGAACTATGAATCTTTTGGAAGTACCAACTAAG 540
ST36            CTTGGCCTTGCTTCAGCAACCTTACTTCGAACTATGAATCTTTTGGAAGTACCAACTAAG 540
ST15            CTTGGCCTTGCTTCAGCAACCTTACTTCGAACCATGAATCTTTTGGAAGTACCAACTAAG 540
                ****************************** ************************** 

ST469           GGACAGATTAGGTTTGAGGGGATTGATATTACCGATAAGAAGAATGATATTTTCAGCATG 600
ST530           GGACAGATTAGGTTTGAGGGGATTGATATTACCGATAAGAAGAATGATATTTTCAGCATG 600
ST36            GGACAGATTAGGTTTGAGGGGATTGATATTACCGATAAGAAGAATGATATTTTCAGCATG 600
ST15            GGACAGATTAGGTTTGAGGGGATTGATATTACCGATAAGAAGAACGATATTTTCAGCATG 600
                ********************************************** *************** 

ST469           CGTGAAAAAATGGGAATGGTTTTCCAGCAGTTTAACCTCTTTCCCAATATGACTATTTTA 660
ST530           CGTGAAAAAATGGGAATGGTTTTCCAGCAGTTTAACCTCTTTCCCAATATGACTATTTTA 660
ST36            CGTGAAAAAATGGGAATGGTTTTCCAGCAGTTTAACCTCTTTCCCAATATGACTATTTTA 660
ST15            CGTGAAAAAATGGGGATGGTTTTTCAGCAGTTTAACCTCTTTCCCAATATGACTGTTTTA 660
                ************** ******** ****************************** ***** 

ST469           GAAAATATCACTTTATCGCCAATCAAAACTAAGGGAATGGCTAAAGCAGAGGCTGACAAA 720
ST530           GAAAATATCACTTTATCGCCAATCAAAACTAAGGGAATGGCTAAAGCAGAGGCTGACAAA 720
ST36            GAAAATATCACTTTATCGCCAATCAAAACTAAGGGAATGGCTAAAGCAGAGGCTGACAAA 720
ST15            GAAAATATCACTTTATCGCCAATCAAAACTAAGGGAATGGCTAAAGCAGAGGCTGACAAA 720
                ************************************************************ 

ST469           ACAGCCTTGAGCTTGTTGGACAAAGTTGGATTATCAGAAAAAGCCAAGGCTTATCCTGCT 780
ST530           ACAGCCTTGAGCTTGTTGGACAAAGTTGGATTATCAGAAAAAGTCAAGGCTTATCCTGCT 780
ST36            ACAGCCTTGAGCTTGTTGGACAAAGTTGGATTATCAGAAAAAGCCAAGGCTTATCCTGCT 780
ST15            ACAGCCTTGAGCTTGTTGGACAAAGTTGGATTATCAGAAAAAGCCAAGGCTTATCCTGCT 780
                ********************************************** ***************** 

ST469           AGTCTTTCTGGTGGGCAACAGCAGCGGATTGCGATTGCGCGTGGACTGGCTATGGATCCA 840
ST530           AGTCTTTCTGGTGGGCAACAGCAGCGGATTGCGATTGCGCGTGGACTAGCTATGGATCCA 840
ST36            AGTCTTTCTGGTGGGCAACAGCAGCGGATTGCGATTGCGCGTGGACTGGCTATGGATCCA 840
ST15            AGTCTTTCTGGTGGGCAACAGCAGCGGATTGCGATTGCGCGTGGACTGGCTATGGATCCA 840
                ********************************************** ************* 

ST469           GATGTTTTACTCTTTGATGAACCAACTTCAGCTCTAGACCCAGAAATGGTGGGTGAGGTC 900
ST530           GATGTTTTACTCTTTGATGAACCAACTTCAGCTCTAGACCCAGAAATGGTGGGTGAGGTC 900
ST36            GATGTTTTACTCTTTGATGAACCAACTTCAGCTCTAGACCCAGAAATGGTGGGTGAGGTC 900
ST15            GATGTTTTACTCTTTGATGAACCAACTTCAGCTCTAGACCCAGAAATGGTGGGTGAGGTC 900
                ************************************************************ 

ST469           TTGGCTGTCATGCAAGATTTGGCTAAATCTGGGATGACTATGGTTATTGCATGCAATACC 960
ST530           TTGGCTGTCATGCAAGATTTGGCTAAATCTGGGATGACTATGGTTATTGCATGCAATACC 960
ST36            TTGGCTGTCATGCAAGATTTGGCTAAATCTGGGATGACTATGGTTATTGCATGCAATACC 960
ST15            TTGGCTGTCATGCAAGATTTGGCTAAATCTGGGATGACTATGGTTATTGCATGCAATACC 960
                ************************************************************ 

ST469           GCAACAGCGGTGGCTTGGGAAGAAGTAAAAGCAGCTTTAGATATTCCTGTTTTAGGGGTT 1020
ST530           GCAACAGCGGTGGCTTGGGAAGAAGTAAAAGCAGCTTTAGATATTCCTGTTTTAGGAGTT 1020
ST36            GCAACAGCGGTGGCTTGGGAAGAAGTAAAAGCAGCTTTAGATATTCCTGTTTTAGGGGTT 1020
ST15            GCAACAGCGGTGGCTTGGGAAGAAGTAAAAGCAGCTTTAGATATTCCTGTTTTAGGGGTT 1020
                ************************************************************ *** 
```

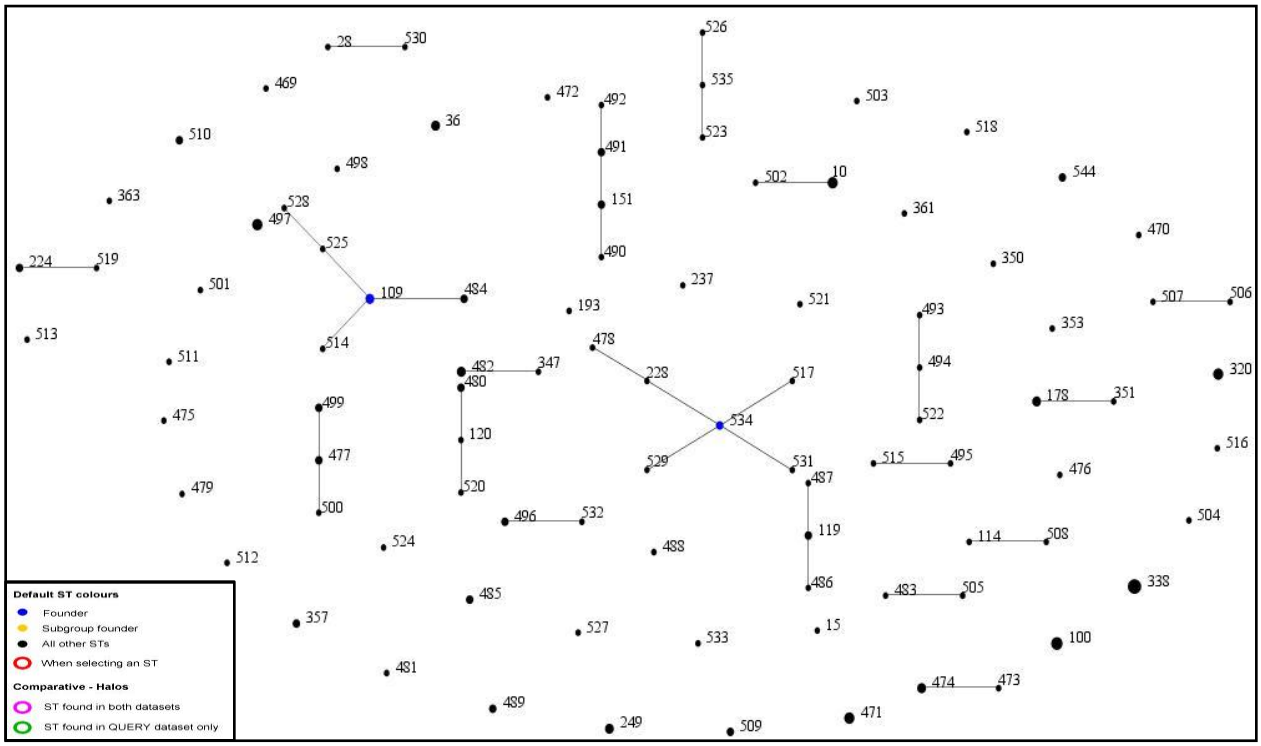
```
ST469           GTCTTACCGGGGGCAAGCGCAGCTATTAAATCAACGACAAAAGGCCAGGTTGGGGTCATC 1080
ST530           GTCTTACCGGGGGCAAGCGCAGCTATTAAATCAACGACAAAAGGCCAGGTTGGGGTCATC 1080
ST36            GTCTTACCGGGGGCAAGCGCAGCTATTAAATCAACGACAAAAGGCCAGGTTGGGGTCATC 1080
ST15            GTCTTACCGGGGGCAAGCGCAGCTATTAAATCAACGACAAAAGGCCAGGTTGGGGTCATC 1080
                ************************************************************

ST469           GGAACCCCAATGACAGTGGCTTCAGACATTTATCGCAAAAAAATCCAGCTATTAGCACCA 1140
ST530           GGAACCCCAATGACAGTGGCTTCAGACATTTATCGCAAAAAAATCCAGCTATTAGCACCA 1140
ST36            GGAACCCCAATGACAGTGGCTTCAGACATTTATCGCAAAAAAATCCAGCTATTAGCACCA 1140
ST15            GGAACCCCAATGACAGTGGCTTCAGACATTTATCGCAAAAAAATCCAGCTATTAGCACCA 1140
                ************************************************************

ST469           TCTATTCAAGTAAGGAGTCTTGCTTGCCCGAAGTTTGTACCGATTGTGGAATCAAATGAG 1200
ST530           TCTATTCAAGTAAGGAGTCTTGCTTGCCCGAAGTTTGTACCGATTGTGGAATCAAATGAG 1200
ST36            TCTATTCAAGTAAGGAGTCTTGCTTGCCCGAAGTTTGTACCGATTGTGGAATCAAATGAG 1200
ST15            TCTGTTCAAGTAAGGAGTCTTGCTTGCCCGAAGTTTGTACCGATTGTGGAATCAAATGAG 1200
                *** ********************************************************

ST469           ATGTGTTCGAGTATAGCTAAAAAAATAGTTTATGACAGTCTAGCACCATTAGTCGGTAAA 1260
ST530           ATGTGTTCGAGTATAGCTAAAAAAATAGTTTATGACAGTCTAGCACCATTAGTCGGTAAA 1260
ST36            ATGTGTTCGAGTATAGCTAAAAAAATAGTTTATGACAGTCTAGCACCATTAGTCGGTAAA 1260
ST15            ATGTGTTCGAGTATAGCTAAAAAAATAGTTTATGACAGTCTAGCACCATTAGTCGGTAAA 1260
                ************************************************************

ST469           ATAGATACCCTTGTACTAGGATGTACTCACTATCCCTTGTTACGACCAATTATCCAAAAT 1320
ST530           ATAGATACCCTTGTACTAGGATGTACTCACTATCCCTTGTTACGACCAATTATCCAAAAT 1320
ST36            ATAGATACCCTTGTACTAGGATGTACTCACTATCCCTTGTTACGACCAATTATCCAAAAT 1320
ST15            ATAGATACCCTTGTACTAGGATGTACTCACTATCCCTTGTTACGACCAATTATCCAAAAT 1320
                ************************************************************

ST469           GTTATGGGGCCATCTGTTAAGCTGATTGACAGTGGAGCAGAATGCGTCCGAGATATCTCT 1380
ST530           GTTATGGGGCCATCTGTTAAGCTGATTGACAGTGGAGCAGAATGCGTCCGAGATATCTCT 1380
ST36            GTTATGGGGCCATCTGTTAAGCTGATTGACAGTGGAGCAGAATGCGTCCGAGATATCTCT 1380
ST15            GTTATGGGGCCATCTGTTAAGCTGATTGACAGTGGAGCAGAATGCGTCCGAGATATCTCT 1380
                ************************************************************

ST469           GTCTTAACCTATATTAATCGTTTACAGAAACTGGCTAAAATTTTGGCAACGGTAGATGTT 1440
ST530           GTCTTAACCTATATTAATCGTTTACAGAAACTGGCTAAAATTTTGGCAACGGTAGATGTT 1440
ST36            GTCTTAACCTATATTAATCGTTTACAGAAACTGGCTAAAACTTTGGCAACGGTGGATGTT 1440
ST15            GTCTTAACCTATATTAATCGTTTACAGAAACTGGCTAAAATTTTGGCAACGGTAGATGTT 1440
                ****************************************** ************ ******

ST469           TTGCAAAGTTTAGCAGTCGTTGCTGAAACCAATCATTATATCCGGCCGCAGTTCAATGAT 1500
ST530           TTGCAAAGTTTAGCAGTCGTTGCTGAAACCAATCATTATATCCGGCCGCAGTTCAATGAT 1500
ST36            TTGCAAAGTTTAGCAGTCGTTGCTGAAACCAATCATTATATTCGGCCGCAGTTCAATGAT 1500
ST15            TTGCAAAGTTTAGCAGTCGTTGCTGAAACCAATCATTATATCCGGCCGCAGTTCAATGAT 1500
                ****************************************** ******************

ST469           AATCATGTGATTACAATTCAAGAAGGTCGTCACGCGGTTGTTGAAAAGGTTATGGGAGTG 1560
ST530           AATCATGTGATTACAATTCAAGAAGGTCGTCACGCGGTTGTTGAAAAGGTTATGGGAGTG 1560
ST36            AATCATGTGATTACAATTCAAGAAGGTCGTCACGCCGTTGTTGAAAAGGTTATGGGAGTG 1560
ST15            AATCATGTGATTACAATTCAAGAAGGTCGTCACGCGGTTGTTGAAAAGGTTATGGGAGTG 1560
                ********************************** ************************
```

```
ST469           CAGGAATACATTCCCAATAGTATCTCTTTTGACCAACAGACCAGTATTCAGCTGATTACA 1620
ST530           CAGGAATACATTCCCAATAGTATCTCTTTTGACCAACAGACCAGTATTCAGCTGATTACA 1620
ST36            CAGGAATACATTCCCAATAGTATCTCTTTTGACCAACAGACCAGTATTCAGCTGATTACA 1620
ST15            CAGGAATACATTCCCAATAGTATCTCTTTTGACCAACAGACCAGTATTCAGCTGATTACA 1620
                ************************************************************

ST469           GGTCCAAATATGAGTGGTAAGTCGACTTATATGAGACAGCTGGCCTTAACGGTTATCATG 1680
ST530           GGTCCAAATATGAGTGGTAAGTCGACTTATATGAGACAGCTGGCCTTAACGGTTATCATG 1680
ST36            GGTCCAAATATGAGTGGTAAGTCGACTTATATGAGACAGCTGGCCTTAACGGTTATCATG 1680
ST15            GGTCCAAATATGAGTGGTAAGTCGACTTATATGAGACAGCTGGCCTTAACGGTTATCATG 1680
                ************************************************************

ST469           GCCCAGATGGGTTCATTTGTGGCTGCTGACCATGTTGATTTACCTTTATTTGATGCGATT 1740
ST530           GCCCAGATGGGTTCATTTGTGGCTGCTGATCATGTTGATTTACCTTTATTTGATGCGATT 1740
ST36            GCCCAGATGGGTTCATTTGTGGCTGCTGATCATGTTGATTTACCTTTATTTGATGCGATT 1740
ST15            GCCCAGATGGGTTCATTTGTGGCTGCTGACCATGTTGATTTACCTTTATTTGATGCGATT 1740
                ***************************** ******************************

ST469           TTTACGCGTATTGGGGCTGCTGATGATTTGATTTCTGGGCAATCAACCTTTCCTATGGCC 1800
ST530           TTTACGCGTATTGGGGCTGCTGATGATTTGATTTCTGGGCAATCAACCTTTCCTATGGCC 1800
ST36            TTTACGCGTATTGGGGCTGCTGATGATTTGATTTCTGGGCAATCAACCTTTCCTATGGCC 1800
ST15            TTTACGCGTATTGGGGCTGCTGATGATTTGATTTCTGGGCAATCAACCTTTCCTATGGCC 1800
                ************************************************************

ST469           TATGTTCTTTGGAATCACTTCATGAACATCAATCCCAAAACAAGCCGTAATTGGTCAAAC 1860
ST530           TATGTTCTTTGGAATCACTTCATGAACATCAATCCCAAAACAAGCCGTAATTGGTCAAAC 1860
ST36            TATGTTCTTTGGAATCACTTCATGAACATCAATCCCAAAACAAGCCGTAATTGGTCAAAC 1860
ST15            TATGTTCTTTGGAATCACTTCATGAACATCAATCCCAAAACAAGCCGTAATTGGTCAAAC 1860
                ************************************************************

ST469           AGAGACCGTTTTATCCTATCAGCAGGTCATGGAAGCGCCATGCTTTATAGCTTGTTGCAT 1920
ST530           AGAGACCGTTTTATCCTATCAGCAGGTCATGGAAGTGCCATGCTTTATAGCTTGTTACAC 1920
ST36            AGAGACCGTTTTATCCTATCAGCAGGTCATGGAAGTGCCATGCTTTATAGCTTGTTGCAT 1920
ST15            AGAGACCGTTTTATCCTATCAGCAGGTCATGGAAGTGCCATGCTTTATAGCTTGTTACAC 1920
                ********************************** ******************** **

ST469           TTAGCAGGTTATGATTTATCTGTAGAAGATTTAAAGAACTTCCGTCAATGGGGTTCTAAA 1980
ST530           TTAGCTGGTTATGATTTATCTGTAGAAGATTTAAAGAACTTCCGTCAATGGGGTTCTAAA 1980
ST36            TTAGCAGGTTATGATTTATCTGTAGAAGATTTAAAGAACTTCCGTCAATGGGCTTCTAAA 1980
ST15            TTAGCTGGTTATGATTTATCTGTAGAAGATTTAAAGAACTTCCGTCAATGGGGTTCTAAA 1980
                ***** ********************************************* *******

ST469           ACACCAGGTCACCCAGAAGTGAACCACACAGACGGTGTCGAAGCAACCACAGGACCTCTT 2040
ST530           ACACCAGGTCACCCAGAAGTGAACCACACAGACGGTGTCGAAGCAACCACAGGACCTCTT 2040
ST36            ACACCAGGTCACCCAGAAGTGAACCACACAGACGGTGTCGAAGCAACCACAGGACCTCTT 2040
ST15            ACACCAGGTCACCCAGAAGTGAACCACACAGACGGTGTCGAAGCAACCACAGGACCTCTT 2040
                ************************************************************

ST469           GGTCAAGGGATCGCAAATGCCGTTGGTATGGCCATGGCAGAAGCTCACCTAGCAGCTAAA 2100
ST530           GGTCAAGGGATCGCAAATGCCGTTGGGATGGCCATGGCAGAAGCTCATCTAGCAGCTAAA 2100
ST36            GGTCAAGGGATCGCAAATGCCGTTGGTATGGCCATGGCAGAAGCTCACCTAGCAGCTAAA 2100
ST15            GGTCAAGGGATCGCAAATGCCGTTGGTATGGCCATGGCAGAAGCTCATCTAGCAGCTAAA 2100
                ************************** ******************** ************
```

```
ST469           TTTAACAAACCAGGCTTTGACATCGTTGATCACTACACATTTGCTTTGAATGGTGACGGT 2160
ST530           TTTAACAAACCAGGCTTTGACATCGTTGATCACTACACATTTGCTTTGAATGGTGACGGT 2160
ST36            TTTAACAAACCAGGCTTTGACATCGTTGATCACTACACATTTGCTTTGAATGGTGACGGT 2160
ST15            TTTAACAAACCAGGCTTTGACATCGTTGATCACTACACCTTTGCTTTGAATGGTGACGGT 2160
                ************************************ ********************

ST469           GACCTTATGGAAGGGGTCAGCCAAGAAGCAGCAAGTATGGCAGGACATTTAAAACTTGGG 2220
ST530           GACCTTATGGAAGGGGTCAGCCAAGAAGCAGCAAGTATGGCAGGACATTTAAAACTTGGG 2220
ST36            GACCTTATGGAAGGGGTCAGCCAAGAAGCAGCAAGTATGGCAGGACATTTAAAACTTGGG 2220
ST15            GACCTTATGGAAGGGGTCAGCCAAGAAGCAGCAAGTATGGCAGGACATTTAAAACTTGGG 2220
                ************************************************************

ST469           AAATTGGTCTTGCTATATGATTCAAACGACGGAGAGAATATTCTAAAGGTAGATAATTTT 2280
ST530           AAATTGGTCTTGCTATATGATTCAAACGACGGAGAGAATATTCTAAAGGTAGATAATTTT 2280
ST36            AAATTGGTCTTGCTATATGATTCAAACGACGGAGAGAATATTCTAAAGGTAGATAATTTT 2280
ST15            AAATTGGTCTTGCTATATGATTCAAACGACGGAGAGAATATTCTAAAGGTAGATAATTTT 2280
                ************************************************************

ST469           TTAACTCATCAAGTTGATTACCGGTTGATGAAAGCAATTGGTAAAGTGTTTGCTCAAAAA 2340
ST530           TTAACTCATCAAGTTGATTACCGGTTGATGAAAGCAATTGGTAAAGTGTTTGCTCAAAAA 2340
ST36            TTAACTCATCAAGTTGATTACCGGTTGATGAAAGCAATTGGTAAAGTGTTTGCTCAAAAA 2340
ST15            TTAACTCATCAAGTTGATTACCGGTTGATGAAAGCAATTGGTAAAGTGTTTGCTCAAAAA 2340
                ************************************************************

ST469           TATGCTGAGGCTGGCATTACAAAAGTGGTTACAATCGAAGCTTCAGGTATTGCACCAGCC 2400
ST530           TATGCTGAGGCTGGCATTACAAAAGTGGTTACAATCGAAGCTTCAGGTATTGCACCAGCC 2400
ST36            TATGCTGAGGCTGGCATTACAAAAGTGGTTACAATCGAAGCTTCAGGTATTGCACCAGCC 2400
ST15            TATGCTGAGGCTGGCATTACAAAAGTGGTTACAATCGAAGCTTCAGGTATTGCACCAGCC 2400
                ************************************************************

ST469           GTATACGCTGCAGAAGCAATGGATGTTCCTATGATTTTTGCGAAAAAACATAAAAACATT 2460
ST530           GTATACGCTGCAGAAGCAATGGATGTTCCTATGATTTTTGCGAAAAAACATAAAAACATT 2460
ST36            GTATACGCTGCAGAAGCAATGGATGTTCCTATGATTTTTGCGAAAAAACATAAAAACATT 2460
ST15            GTATACGCTGCAGAAGCAATGGATGTTCCTATGATTTTTGCGAAAAAACATAAAAACATT 2460
                ************************************************************

ST469           ACCATGACAGAAGGCATTTTGACAGCAGAAGTTTATTCTTTCACTAAACAAGTGACGAGC 2520
ST530           ACCATGACAGAAGGCATTTTGACAGCAGAAGTTTATTCTTTCACTAAACAAGTGACGAGC 2520
ST36            ACCATGACAGAAGGCATTTTGACAGCAGAAGTTTATTCTTTCACTAAACAAGTGACGAGC 2520
ST15            ACCATGACAGAAGGCATTTTGACAGCAGAAGTTTATTCTTTCACTAAACAAGTGACGAGC 2520
                ************************************************************

ST469           ACGGTGTCTATCGCTGGTAAATTCCTATCTAAAGAAGACAAGGTTTTGATTATTGATGAC 2580
ST530           ACGGTGTCTATCGCTGGTAAATTCCTATCTAAAGAAGACAAGGTTTTGATTATTGATGAC 2580
ST36            ACGGTGTCTATCGCTGGTAAATTCCTATCTAAAGAAGACAAGGTTTTGATTATTGATGAC 2580
ST15            ACGGTGTCTATCGCTGGTAAATTCCTATCTAAAGAAGACAAGGTTTTGATTATTGATGAC 2580
                ************************************************************

ST469           TTTTTAGCTAATGGTCAGGCAGCCAAAGGCTTGATCGAGATTATTGGTCAAGCAGGGGCA 2640
ST530           TTTTTAGCTAATGGTCAGGCAGCCAAAGGCTTGATTGAGATTATTGGTCAAGCAGGGGCA 2640
ST36            TTTTTAGCTAATGGTCAGGCAGCCAAAGGCTTGATTGAGATTATTGGTCAAGCAGGGGCA 2640
ST15            TTTTTAGCTAATGGTCAGGCAGCCAAAGGCTTGATCGAGATTATTGGTCAAGCAGGGGCA 2640
                ********************************** ************************
```

```
ST469           CAAGTCGTCGGCGTTGGTATTGTGATTGAGAAATCTTTCCAAGATGGTCGTCGATTGATT 2700
ST530           CAAGTCGTCGGCGTTGGTATTGTGATTGAGAAATCTTTCCAAGATGGTCGTCGATTGATT 2700
ST36            CAAGTCGTCGGCGTTGGTATTGTGATTGAGAAATCTTTCCAAGATGGTCGTCGATTGATT 2700
ST15            CAAGTCGTCGGCGTTGGTATTGTGATTGAGAAATCTTTCCAAGATGGTCGTCGATTGATT 2700
                ************************************************************

ST469           TTTGCAGCTGATAGCCAACGTAAAGCCCAACTTGCCATAGAAAAAGGTCGTTTCAAAGAA 2760
ST530           TTTGCAGCTGATAGCCAACGTAAAGCCCAACTTGCCATAGAAAAAGGTCGTTTCAAAGAA 2760
ST36            TTTGCAGCTGATAGCCAACGTAAAGCCCAACTTGCCATAGAAAAAGGTCGTTTCAAAGAA 2760
ST15            TTTGCAGCTGATAGCCAACGTAAAGCCCAACTTGCCATAGAAAAAGGTCGTTTCAAAGAA 2760
                ************************************************************

ST469           GAGATTGCACCTGTCACTATTCCTCAGCGTAAAGGTGAACCTTTACTCGTTGATCAAGAT 2820
ST530           GAGATTGCACCTGTCACTATTCCTCAGCGTAAAGGTGAACCTTTACTCGTTGATCAAGAT 2820
ST36            GAGATTGCACCTGTCACTATTCCTCAGCGTAAAGGTGAACCTTTACTCGTTGATCAAGAT 2820
ST15            GAGATTGCACCTGTCACTATTCCTCAGCGTAAAGGTGAACCTTTACTCGTTGATCAAGAT 2820
                ************************************************************

ST469           GAATACCCTAAATTTGGAACGACAGTGGATAAGTTAGCAAAGTTACGCCCTGCTTTTATC 2880
ST530           GAATACCCTAAATTTGGAACGACAGTGGATAAGTTAGCAAAGTTACGCCCTGCTTTTATC 2880
ST36            GAATACCCTAAATTTGGAACGACAGTGGATAAGTTAGCAAAGTTACGCTCTGCTTTTATC 2880
ST15            GAATACCCTAAATTTGGAACGACAGTGGATAAGTTAGCAAAGTTACGCTCTGCTTTTATC 2880
                ********************************************** ***********

ST469           AAAGATGAGGGGACAGTAACTGCTGGTAATGCTTCAGGAATCAATGATGGAGCAGCAGCA 2940
ST530           AAAGATGAGGGGACAGTAACTGCTGGTAATGCTTCAGGAATCAATGATGGAGCAGCGGCA 2940
ST36            AAAGATGAGGGGACAGTAACTGCTGGTAATGCTTCAGGAATCAATGATGGAGCAGCGGCA 2940
ST15            AAAGATGAGGGGACAGTAACTGCTGGTAATGCTTCAGGAATCAATGATGGAGCAGCGGCA 2940
                ************************************************************ ***

ST469           ATTTTATTGATGAGTAAAGAAAAAGCTGAAGAATTAGGGCTCCCTATTTTAGCTAAAATC 3000
ST530           ATTTTATTGATGAGTAAAGAAAAAGCTGAAGAATTAGGGCTCCCTATTTTAGCTAAAATC 3000
ST36            ATTTTATTGATGAGTAAAGAAAAAGCTGAAGAATTAGGGCTCCCTATTTTAGCTAAAATC 3000
ST15            ATTTTATTGATGAGTAAAGAAAAAGCTGAAGAATTAGGGCTCCCTATTTTAGCTAAAATC 3000
                ************************************************************

ST469           ACTAGTTATGCAAGTGCAGGTGTAGACCCAAGTATTATGGGCTGCGGACCAATACCTGCT 3060
ST530           ACTAGTTATGCAAGTGCAGGTGTAGACCCAAGTATTATGGGCTGCGGACCAATACCTGCT 3060
ST36            ACTAGTTATGCAAGTGCAGGTGTAGACCCAAGTATTATGGGCTGCGGACCAATACCTGCT 3060
ST15            ACTAGTTATGCAAGTGCAGGTGTAGACCCAAGTATTATGGGCTGCGGACCAATACCTGCT 3060
                ************************************************************

ST469           ACGAAAAAGGCTCTTGCAAAGGCTCAGCTGACAATTGATGACATTGATTTGATTGAAGCA 3120
ST530           ACGAAAAAGGCTCTTGCAAAGGCTCAGCTGACAATTGATGACATTGATTTGATTGAAGCA 3120
ST36            ACGAAAAAGGCTCTTGCAAAGGCTCAGCTGACAATTGATGACATTGATTTGATTGAAGCA 3120
ST15            ACGAAAAAGGCTCTTGCAAAGGCTCAGCTGACAATTGATGACATTGATTTGATTGAAGCA 3120
                ************************************************************

ST469           AACGAAGCTTTTGC 3134
ST530           AACGAAGCTTTTGC 3134
ST36            AACGAAGCTTTTGC 3134
ST15            AACGAAGCTTTTGC 3134
                **************
```

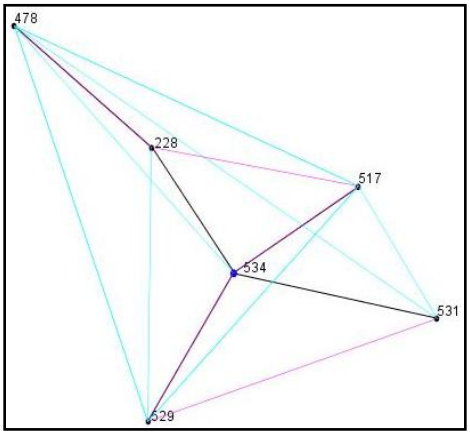
Fig.8. eBURST Snap shot of Community and Hospital GAS isolates (*n = 143*)



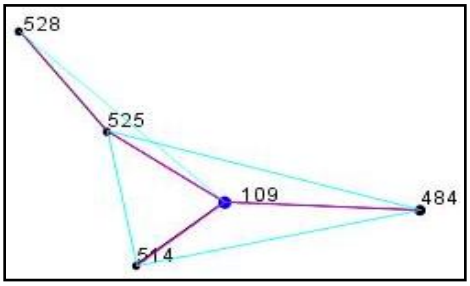
Black line represents groups created by eBurst definition, Pink line represents SLV, Blue line represents DLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder.

Fig.9.Clonal Complexes found among the 93 ST's from Hospital and Community GAS Isolates

a. Group 1 Clonal Complex

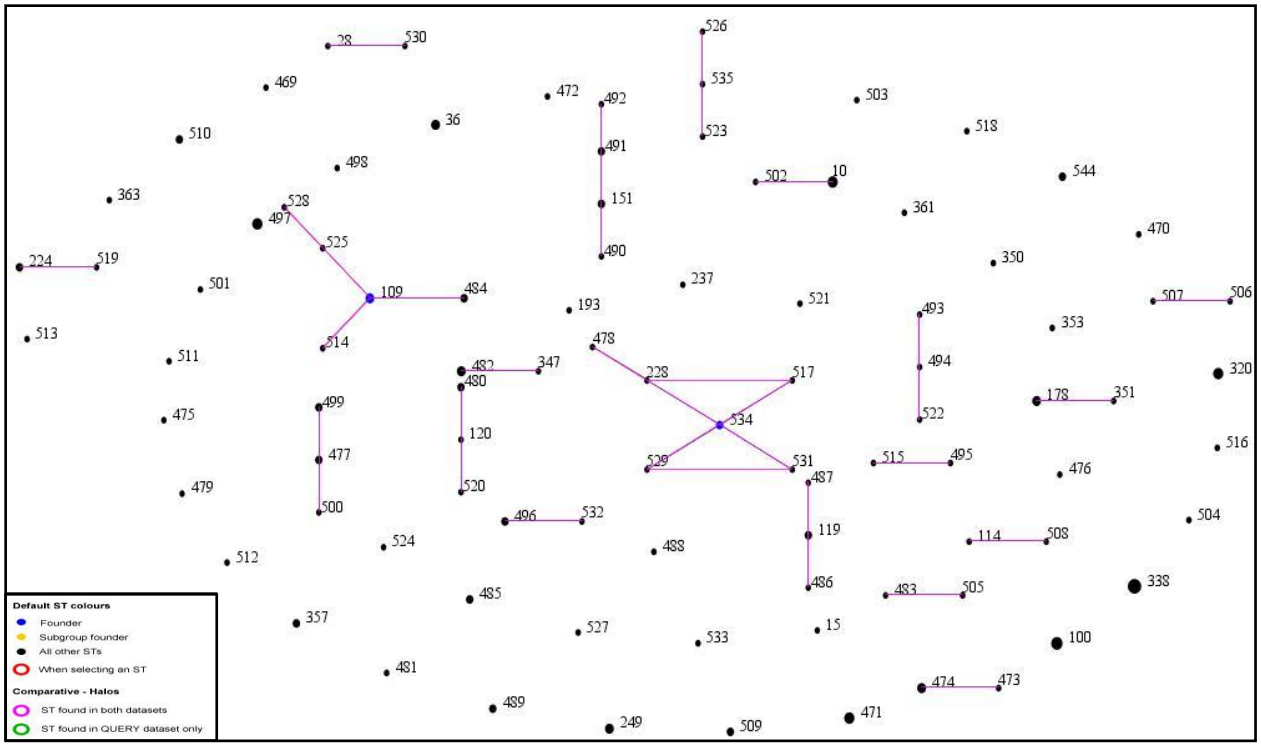


b.Group 2 Clonal Complex



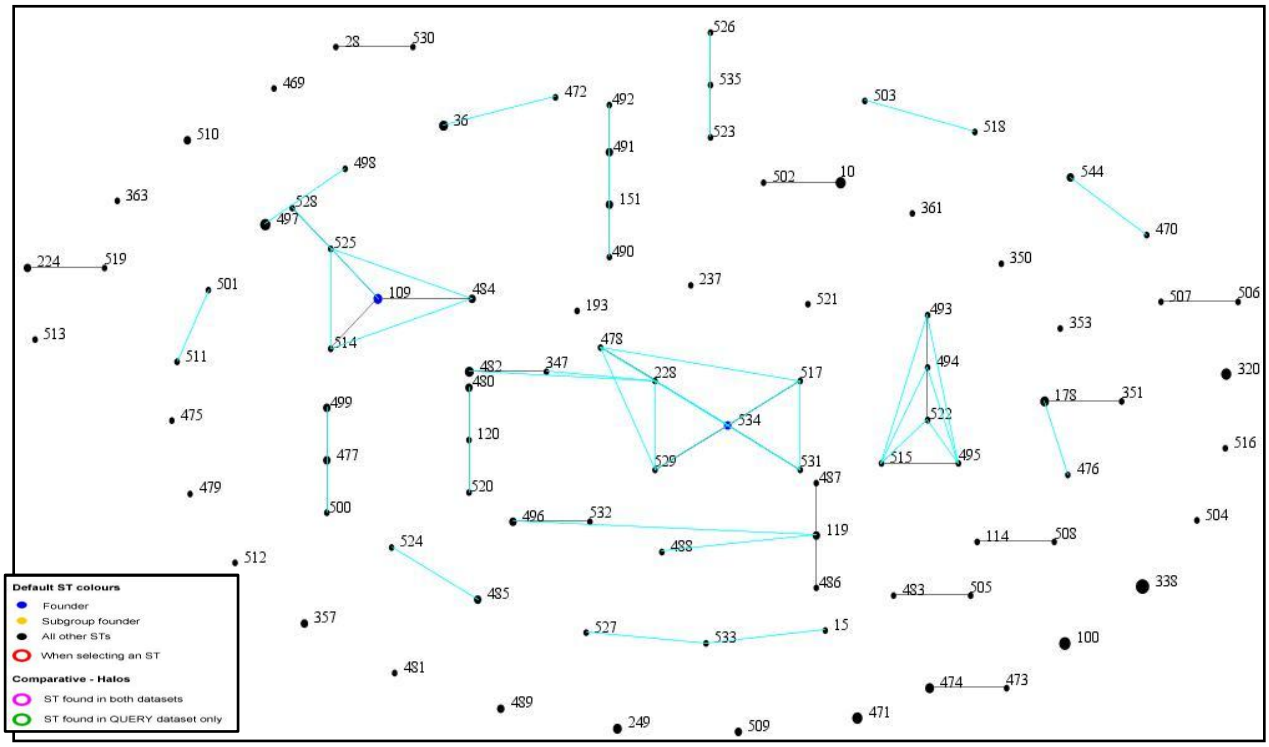
Black line represents groups created by eBurst definition, Pink line represents SLV, Blue line represents DLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder. Note: the clonal complex diagram has been slightly modified from the original population snap shot so the SLV and the DLV can be appreciated in the figure.

Fig.10.eBURST Snap shot of Community and Hospital GAS isolates with Single locus Variant (*n = 143*)



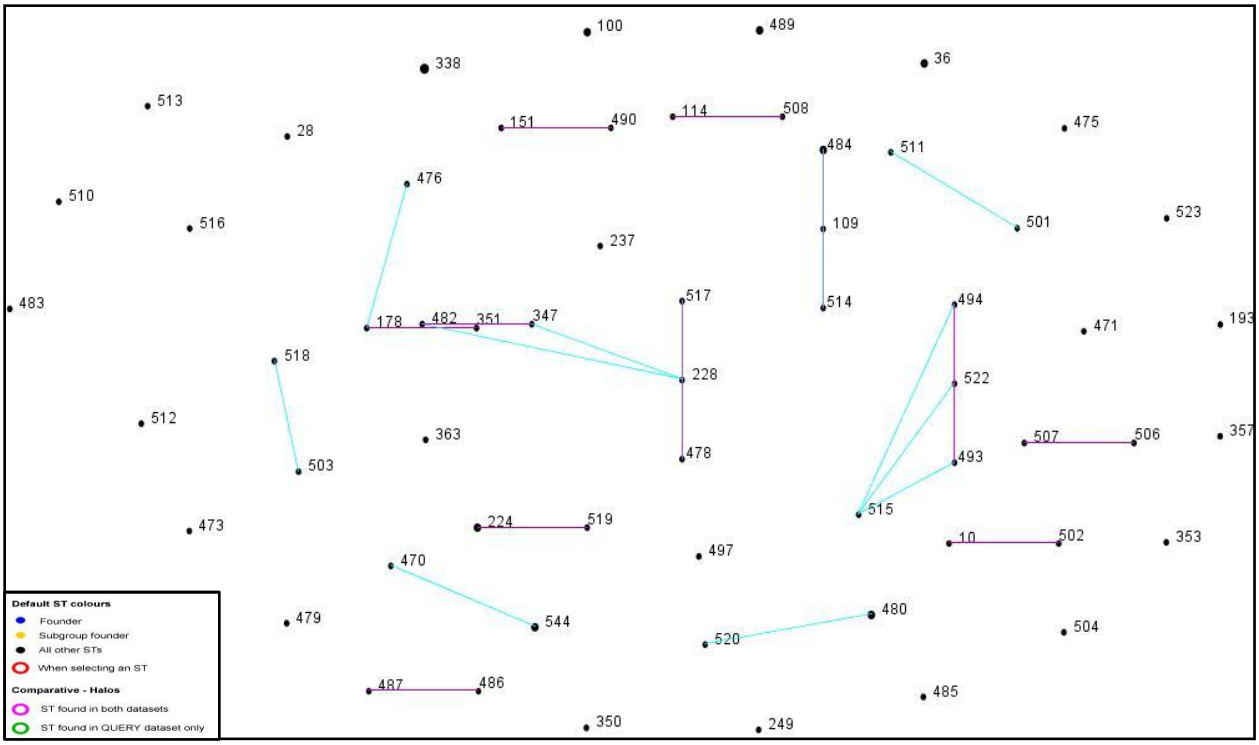
Pink line represents SLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder.

Fig.11.eBURST Snap shot of Community and Hospital GAS isolates with Double locus Variant (*n = 143*)



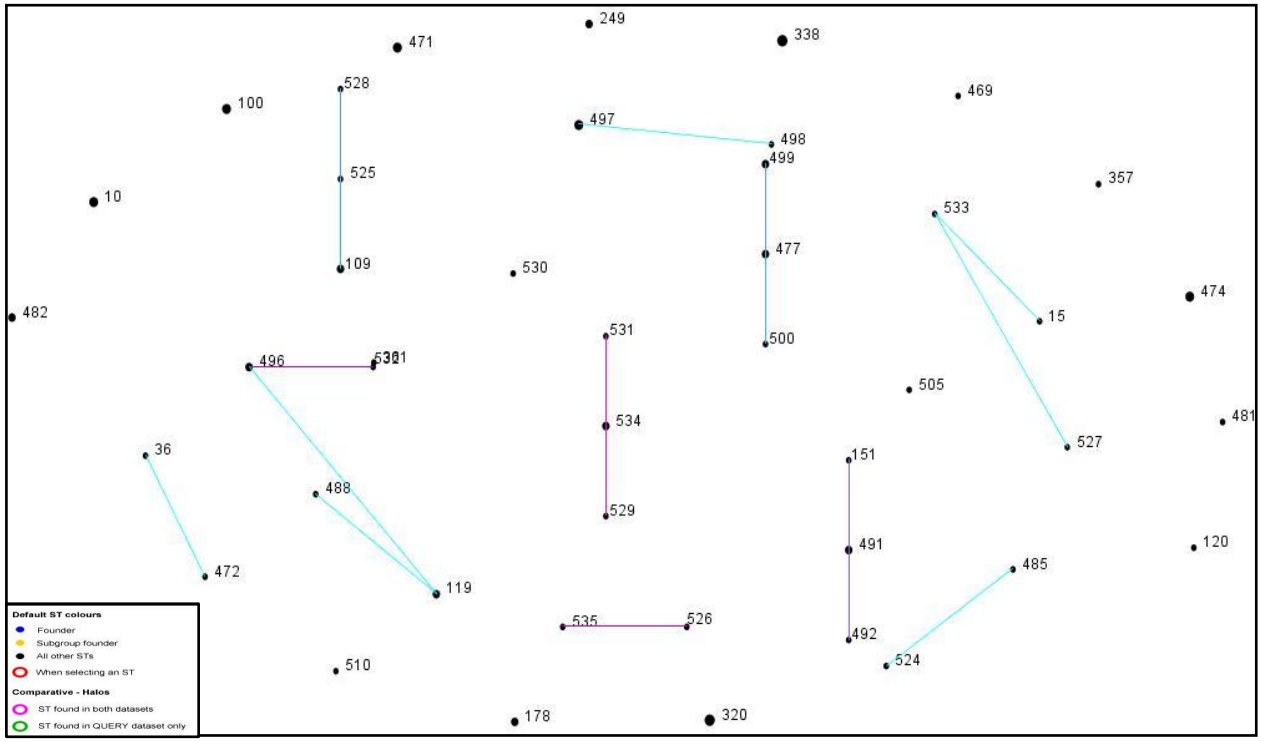
Black line represents groups created by eBurst definition, Blue line represents DLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder

Fig.12.eBURST Snap shot of Hospital GAS isolates (*n = 70*)



Black line represents groups created by eBurst definition, Pink line represents SLV, Blue line represents DLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder.

Fig.13.eBURST Snap shot of Community GAS isolates (*n = 73*)



Black line represents groups created by eBurst definition, Pink line represents SLV, Blue line represents DLV, Black dot represents the STs and size of the dot represents the number of GAS isolate in each ST, Blue dot represents the predicted founder.

Fig.14.Phylogenetic Tree of *emm* types

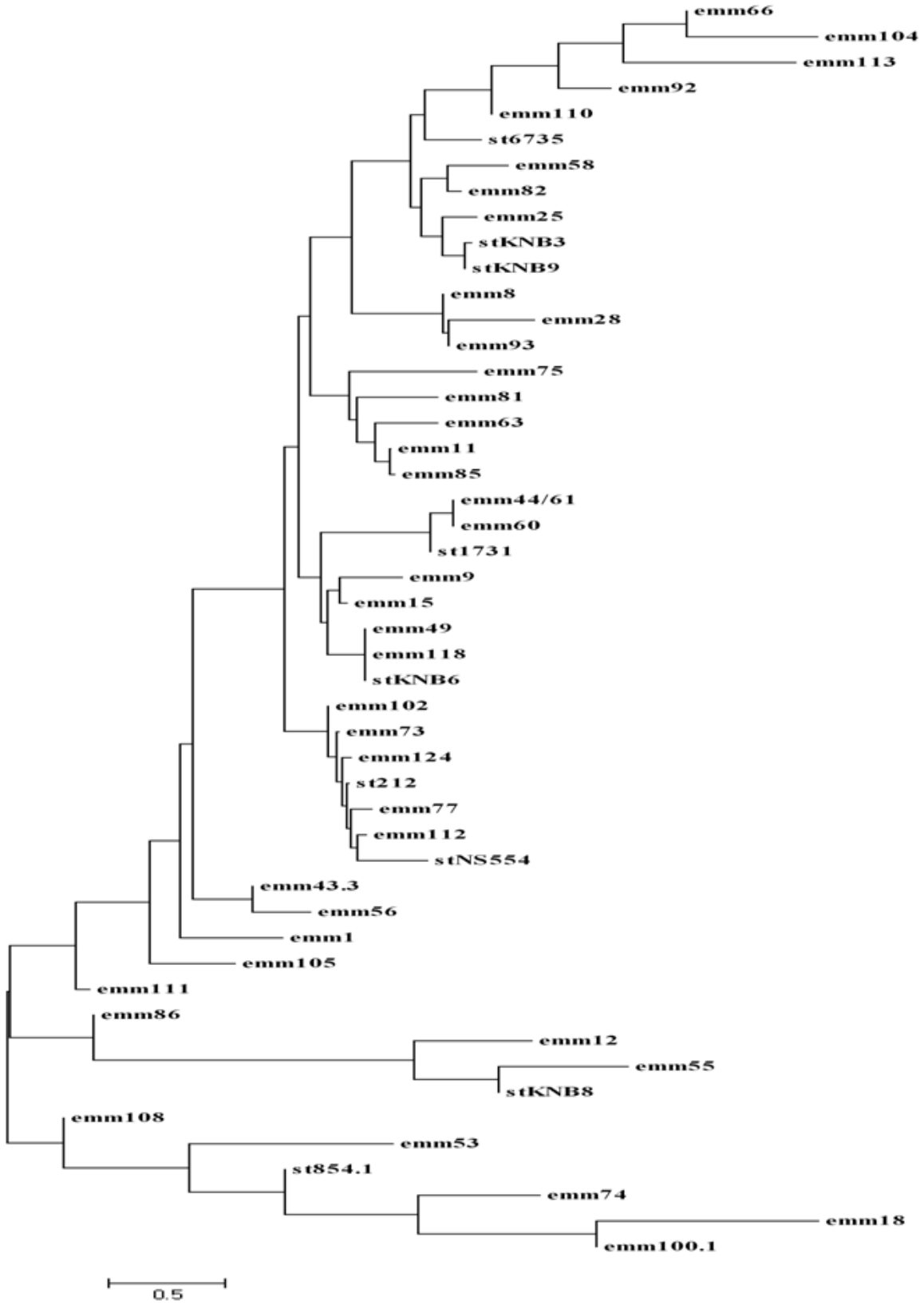


Fig.15.Phylogenetic Tree of Sequence types

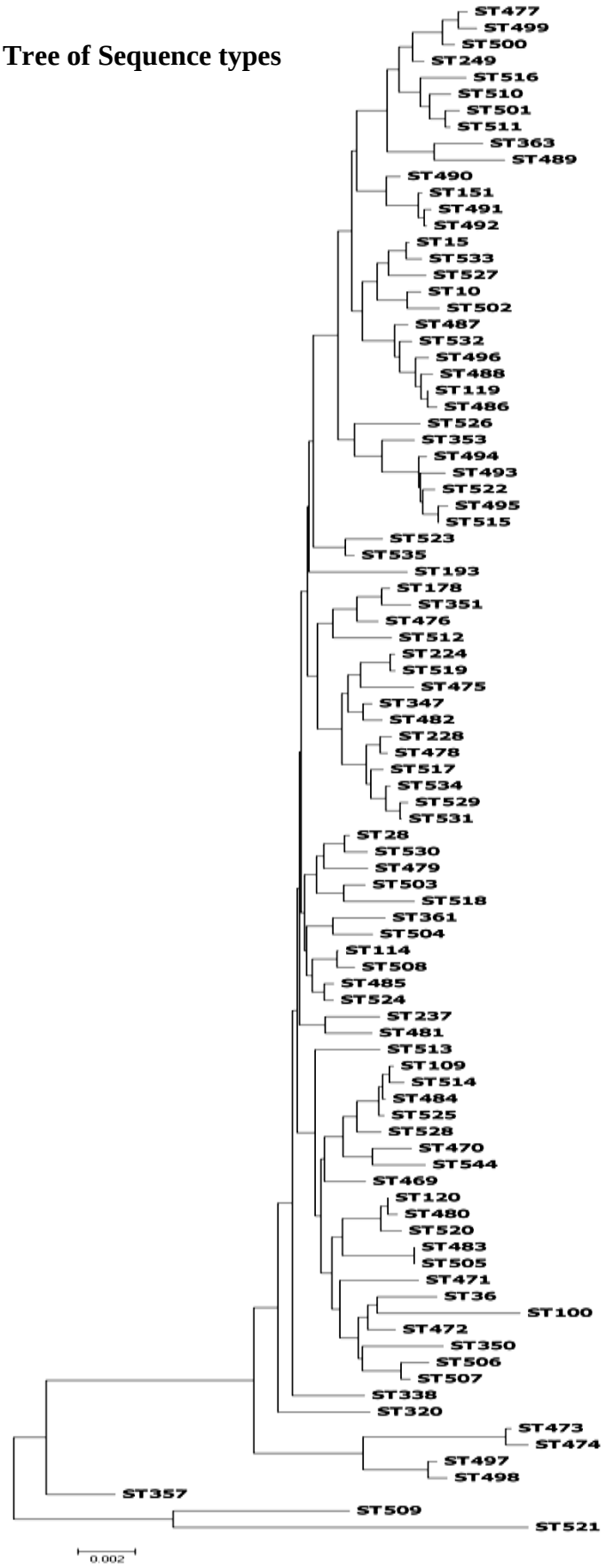


Table.1. Distribution of *emm* types among the community GAS isolates *(n=319)*

emm type	Pharyngitis	Normal Throat	Impetigo	Normal skin	Total
112	1	18	8	2	29
11	9	4	2	0	15
82	5	4	1	2	12
100	0	3	6	2	11
63	4	3	2	1	10
93	2	0	4	4	10
85	4	3	2	0	9
105	1	5	2	1	9
108	5	2	0	2	9
49.4	2	4	1	1	8
110	8	0	0	0	8
113	2	4	2	0	8
1	3	3	1	0	7
44	0	1	3	3	7
66	0	5	2	0	7
111	0	4	1	2	7
18	3	2	0	1	6
55	3	1	2	0	6
92	3	1	2	0	6
12	2	0	2	1	5
28	1	2	2	0	5
36	0	2	0	3	5
56	0	3	1	1	5
69	1	1	1	2	5
75	0	3	1	1	5
77	2	1	0	2	5
86	2	3	0	0	5
st1389	1	0	2	2	5
st2147	2	2	0	1	5
53	1	2	1	0	4
68	1	2	1	0	4
74	3	0	0	1	4
102	1	0	2	1	4
15	0	2	1	0	3
48	0	0	2	1	3
58	3	0	0	0	3
60	2	1	0	0	3
78	2	1	0	0	3

emm type	Pharyngitis	Normal Throat	Impetigo	Normal skin	Total
104	2	1	0	0	3
st212	2	1	0	0	3
st854	0	2	1	0	3
stNS554	2	1	0	0	3
9	2	0	0	0	2
65	0	2	0	0	2
103	0	2	0	0	2
116	1	0	1	0	2
119	0	1	1	0	2
st2460	2	0	0	0	2
stC36	0	1	1	0	2
stD432	1	0	1	0	2
stKNB3	0	2	0	0	2
stKNB8	1	1	0	0	2
stNS292	2	0	0	0	2
3	1	0	0	0	1
25	0	0	0	1	1
33	1	0	0	0	1
54	0	1	0	0	1
57	0	0	1	0	1
70	0	1	0	0	1
71	0	0	1	0	1
73	0	1	0	0	1
79	0	0	0	1	1
81.2	0	1	0	0	1
88	1	0	0	0	1
122	0	0	1	0	1
st11014	1	0	0	0	1
st1731	1	0	0	0	1
st465	1	0	0	0	1
st5282	0	1	0	0	1
st62	1	0	0	0	1
stCQ343	0	0	0	1	1
stG245	0	1	0	0	1
stG485	0	1	0	0	1
Total	**101**	**113**	**65**	**40**	**319**

Table.2.Distribution of *emm* types common to Community and hospital associated Pharyngitis

emm* type	**Community Pharyngitis (101)**	**Hospital Pharyngitis (41)**
11	9 (8.9%)	1(2.4%)
110	8(7.9%)	-
82	5(4.9%)	-
108	5(4.9%)	-
63	4(3.96%)	1(2.4%)
85	4(3.96%)	1(2.4%)
1	3(2.97%)	1(2.4%)
18	3(2.97%)	-
55	3(2.97%)	3(7.3%)
92	3(2.97%)	-
74	3(2.97%)	-
58	3(2.97%)	-

*14 types that accounted for 52.5% of the 101 GAS isolates

Table.3.Distribution of Invasive *emm* types among Community and Hospital Pharyngeal GAS isolates

emm* Type	**Invasive GAS Disease (60)**	**Community Associated Pharyngitis(101)**	**Hospital Associated Pharyngitis (41)**
56	2	-	-
58	2	3	-
63	3	4	1
66	2	-	-
74	3	3	-
85	3	4	1
86	2	2	-
93	2	2	2
100	2	-	-
102	2	1	-
105	2	1	-
110	4	8	-
124	2	-	-
st6735	2	-	-
st854.1	3	-	-

*Represent 36 (62%) of the 58 GAS isolates associated with invasive disease; remaining 24 isolates accounted for one isolate each.

Table.4.GAS isolates selected for MLST from the community (n=73)

emm Type	Pharyngitis	Normal Throat	Impetigo	Normal Skin	Number of strains selected
1	1	1	1	0	3
11	1	1	1	0	3
12	1	0	1	0	2
18	1	1	0	1	3
28	1	1	1	0	3
44	0	1	1	1	3
49	1	1	1	1	4
55	1	1	1	0	3
63	1	1	1	1	4
66	0	1	1	0	2
75	0	0	0	1	1
77	1	1	0	0	2
82	1	1	1	1	4
85	1	1	1	0	3
92	1	1	1	0	3
93	1	0	1	1	3
100	0	1	1	1	3
105	1	1	1	1	4
108	1	1	0	1	3
110	1	0	0	0	1
111	0	1	1	1	3
112	1	1	1	1	4
113	2	1	1	0	4
st212	0	1	0	0	1
stKNB3	0	1	0	0	1
stKNB8	1	1	0	0	2
stNS554	1	0	0	0	1
Total	**21**	**22**	**18**	**12**	**73**

Table.5. STs of GAS isolates from the community and the hospital *(n=143)*

gki	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm*-types	No.
2	6	8	5	2	3	2	**15**	1	1
26	2	2	5	31	3	1	**469**	1	1
4	25	4	4	4	2	4	**530**	1	1
5	2	2	6	6	2	2	**36**	1,12	3
51	2	8	25	33	3	75	**470**	8	1
3	4	70	7	1	5	4	**471**	11	4
4	3	4	4	4	2	4	**28**	12	1
4	2	2	58	6	2	2	**472**	12	1
47	3	2	5	10	64	12	**513**	15.1	1
43	2	2	7	1	3	1	**120**	18	1
5	2	2	3	2	34	31	**526**	18	1
4	2	2	3	2	34	31	**535**	18	1
4	28	2	3	2	34	31	**523**	18.12	1
14	2	8	6	55	40	6	**350**	25.2	1
4	36	8	7	51	3	76	**473**	28	1
4	79	8	7	51	3	76	**474**	28	3
1	31	8	6	36	53	4	**475**	43.3	1
4	4	3	6	14	3	4	**178**	44	3
24	4	3	6	14	3	4	**351**	44	1
4	3	3	6	14	3	77	**476**	44	1
57	3	8	25	1	3	78	**477**	44,113	2
4	31	8	11	34	3	40	**228**	49.4	1
4	2	8	11	34	3	40	**478**	49.4	1
4	6	8	3	34	3	40	**529**	49.4	1
4	35	8	3	34	3	40	**531**	49.4	1
4	31	8	3	34	3	40	**534**	49.4	2
90	6	8	7	4	3	59	**363**	53	1
4	37	8	2	33	3	4	**528**	55	1
34	2	2	21	1	29	16	**100**	55,92	5
40	2	9	6	41	34	12	**512**	56	1
29	32	2	5	33	5	4	**516**	58.8	1
48	31	41	9	2	2	36	**193**	60	1
85	2	14	5	19	58	58	**338**	63	7
13	31	8	5	19	3	1	**249**	66	3
103	37	2	24	4	2	79	**479**	73	1
43	2	2	7	1	3	80	**480**	74	2
43	32	2	7	1	3	1	**520**	74	1
16	35	2	5	61	3	25	**481**	75	1
75	2	44	4	66	3	1	**357**	75,108	2
4	31	2	11	34	3	2	**347**	77	1
4	31	2	11	34	3	21	**482**	77	3
79	37	44	36	51	8	25	**521**	81.2	1
4	63	8	5	73	3	4	**320**	82	4
42	37	8	22	33	3	4	**109**	85	3
42	2	8	22	33	3	4	**484**	85	2
42	37	8	22	33	3	63	**514**	85	1
4	37	8	22	33	3	4	**525**	85	1
4	2	36	5	4	23	1	**503**	86.2	1
4	12	36	11	4	23	1	**518**	86.2	1
4	3	2	59	2	2	4	**485**	92	2
42	3	2	5	2	2	4	**524**	92	1
2	2	9	13	2	14	2	**10**	93	4
2	6	9	13	2	14	2	**502**	93	1
2	6	8	3	2	32	1	**119**	100	2
2	3	8	3	2	32	1	**487**	100	1
55	6	8	3	90	32	1	**488**	100	1

gki	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	**ST**	***emm*-types**	**No.**
2	12	8	3	2	32	1	**486**	100.3	1
13	45	71	4	76	3	6	**489**	102.2	2
25	37	8	7	2	3	1	**353**	104	1
2	28	2	17	4	5	1	**151**	105	2
2	31	2	17	4	5	1	**490**	105	1
2	28	2	17	4	76	1	**491**	105	2
55	28	2	17	4	76	1	**492**	105	1
2	6	3	22	2	2	2	**527**	108	1
2	6	8	2	2	2	2	**533**	108	1
25	4	8	6	30	25	4	**493**	110	1
25	3	8	6	30	25	4	**494**	110	1
25	81	8	6	30	25	39	**495**	110	1
25	37	8	6	30	25	4	**522**	110	1
25	2	8	6	30	25	39	**515**	110.3	1
2	6	8	3	2	8	22	**496**	111	2
2	31	8	3	2	8	22	**532**	111	1
4	2	2	60	21	3	81	**497**	112	4
4	19	3	60	21	3	81	**498**	112	1
57	3	8	25	1	3	76	**499**	113	2
57	3	8	25	19	3	78	**500**	113	1
29	32	8	5	48	5	25	**501**	118	1
29	31	2	5	48	5	25	**511**	118	1
4	59	2	5	44	33	1	**237**	119.2	1
51	49	8	25	33	3	27	**544**	124	2
51	2	2	25	2	5	31	**504**	st1731	1
104	80	2	7	35	2	2	**483**	st212	1
104	82	2	7	35	2	2	**505**	st212	1
105	6	2	6	1	49	4	**506**	st6735	1
105	3	2	6	1	49	4	**507**	st6736	1
4	3	54	6	34	8	44	**224**	st854.1	2
4	2	54	6	34	8	44	**519**	st854.1	1
89	3	63	25	4	3	32	**361**	stKNB3	1
4	36	12	18	2	3	4	**508**	stKNB3	1
4	31	8	5	34	3	40	**517**	stKNB6	1
106	13	3	61	1	28	82	**509**	stKNB8	2
4	36	2	18	2	3	4	**114**	stKNB9	1
29	31	6	5	19	77	4	**510**	stNS554	2

Glucose kinases (gki), Glutamine transporter protein (gtr), Glutamate racemase (murI), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table.6.Allelic profile of newly identified Sequence Types (ST) *(n=68)*

ST	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*
469	26	2	2	5	31	3	1
470	51	2	8	25	33	3	75
471	3	4	70	7	1	5	4
472	4	2	2	58	6	2	2
473	4	36	8	7	51	3	76
474	4	79	8	7	51	3	76
475	1	31	8	6	36	53	4
476	4	3	3	6	14	3	77
477	57	3	8	25	1	3	78
478	4	2	8	11	34	3	40
479	103	37	2	24	4	2	79
480	43	2	2	7	1	3	80
481	16	35	2	5	61	3	25
482	4	31	2	11	34	3	21
483	104	80	2	7	35	2	2
484	42	2	8	22	33	3	4
485	4	3	2	59	2	2	4
486	2	12	8	3	2	32	1
487	2	3	8	3	2	32	1
488	55	6	8	3	90	32	1
489	13	45	71	4	76	3	6
490	2	31	2	17	4	5	1
491	2	28	2	17	4	76	1
492	55	28	2	17	4	76	1
493	25	4	8	6	30	25	4
494	25	3	8	6	30	25	4
495	25	81	8	6	30	25	39
496	2	6	8	3	2	8	22
497	4	2	2	60	21	3	81
498	4	19	3	60	21	3	81
499	57	3	8	25	1	3	76
500	57	3	8	25	19	3	78
501	29	32	8	5	48	5	25
502	2	6	9	13	2	14	2
503	4	2	36	5	4	23	1
504	51	2	2	25	2	5	31
505	104	82	2	7	35	2	2
506	105	6	2	6	1	49	4
507	105	3	2	6	1	49	4
508	4	36	12	18	2	3	4
509	106	13	3	61	1	28	82
510	29	31	6	5	19	77	4
511	29	31	2	5	48	5	25
512	40	2	9	6	41	34	12
513	47	3	2	5	10	64	12
514	42	37	8	22	33	3	63
515	25	2	8	6	30	25	39
516	29	32	2	5	33	5	4
517	4	31	8	5	34	3	40
518	4	12	36	11	4	23	1
519	4	2	54	6	34	8	44
520	43	32	2	7	1	3	1
521	79	37	44	36	51	8	25
522	25	37	8	6	30	25	4
523	4	28	2	3	2	34	31
524	42	3	2	5	2	2	4

ST	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*
525	4	37	8	22	33	3	4
526	5	2	2	3	2	34	31
527	2	6	3	22	2	2	2
528	4	37	8	2	33	3	4
529	4	6	8	3	34	3	40
530	4	25	4	4	4	2	4
531	4	35	8	3	34	3	40
532	2	31	8	3	2	8	22
533	2	6	8	2	2	2	2
534	4	31	8	3	34	3	40
535	4	2	2	3	2	34	31
544	51	49	8	25	33	3	27

Glucose kinases (gki), Glutamine transporter protein (gtr), Glutamate racemase (murI), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table.7.Allelic profile of GAS Pharyngitis from the community*(n=21)*

gki	***gtr***	***muri***	***muts***	***recp***	***xpt***	***yqil***	**ST**	***emm* type**
26	2	2	5	31	3	1	**469***	**1**
3	4	70	7	1	5	4	**471***	**11**
4	2	2	58	6	2	2	**472***	**12**
5	2	2	3	2	34	31	**526***	**18**
4	79	8	7	51	3	76	**474***	**28**
4	31	8	3	34	3	40	**534***	**49.4**
34	2	2	21	1	29	16	**100**	**55**
85	2	14	5	19	58	58	**338**	**63**
4	31	2	11	34	3	21	**482***	**77**
4	63	8	5	73	3	4	**320**	**82**
4	37	8	22	33	3	4	**525***	**85**
42	3	2	5	2	2	4	**524***	**92**
2	2	9	13	2	14	2	**10**	**93**
2	28	2	17	4	5	1	**151**	**105**
2	6	3	22	2	2	2	**527***	**108**
25	81	8	6	30	25	39	**495***	**110**
4	2	2	60	21	3	81	**497***	**112**
57	3	8	25	1	3	76	**499***	**113**
57	3	8	25	1	3	76	**499***	**113**
106	13	3	61	1	28	82	**509***	**stKNB8**
29	31	6	5	19	77	4	**510***	**stNS554**

Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*)Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*), *=New types

Table.8.Allelic profile of Normal Throat carriers of GAS from community*(n=22)*

gki	***gtr***	***muri***	***muts***	***recp***	***xpt***	***yqil***	**ST**	***emm* type**
4	25	4	4	4	2	4	530*	**1**
3	4	70	7	1	5	4	471*	**11**
43	2	2	7	1	3	1	120	**18**
4	79	8	7	51	3	76	474*	**28**
57	3	8	25	1	3	78	477*	**44**
4	6	8	3	34	3	40	529*	**49.4**
4	37	8	2	33	3	4	528*	**55**
85	2	14	5	19	58	58	338	**63**
13	31	8	5	19	3	1	249	**66**
4	31	2	11	34	3	21	482*	**77**
4	63	8	5	73	3	4	320	**82**
42	37	8	22	33	3	4	109	**85**
4	3	2	59	2	2	4	485*	**92**
2	6	8	3	2	32	1	119	**100**
2	28	2	17	4	76	1	491*	**105**
75	2	44	4	66	3	1	357	**108**
2	6	8	3	2	8	22	496*	**111**
4	2	2	60	21	3	81	497*	**112**
57	3	8	25	19	3	78	500*	**113**
104	82	2	7	35	2	2	505*	**st212**
89	3	63	25	4	3	32	361	**stKNB3**
106	13	3	61	1	28	82	509*	**stKNB8**

Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*) ,Acetyl-coA acetyl transferase (*Yqil*) *=New types

Table.9. Genetic relatedness among *emm* types from community Pharyngitis and Normal throat

***emm* type**	**ST of Pharyngitis GAS**	**ST of Normal Throat GAS**
11	471	471
28	474	474
63	338	338
77	482	482
82	320	320
112	497	497
stKNB8	509	509

Table.10. Genetic diversity among *emm* types from community Pharyngitis and Normal throat

***emm* type**	**ST of Pharyngitis GAS**	**ST of Normal Throat GAS**
1	469	530
18	526	120
49	534	529
55	100	528
85	525	109
92	524	485
105	151	491
113	499	500

Table.11.Allelic profile of GAS Impetigo from community *(n=18)*

gki	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
2	6	8	5	2	3	2	**15**	**1**
3	4	70	7	1	5	4	**471***	**11**
5	2	2	6	6	2	2	**36**	**12**
4	79	8	7	51	3	76	**474***	**28**
4	4	3	6	14	3	4	**178**	**44**
4	35	8	3	34	3	40	**531***	**49.4**
34	2	2	21	1	29	16	**100**	**55**
85	2	14	5	19	58	58	**338**	**63**
13	31	8	5	19	3	1	**249**	**66**
4	63	8	5	73	3	4	**320**	**82**
42	37	8	22	33	3	4	**109**	**85**
34	2	2	21	1	29	16	**100**	**92**
2	2	9	13	2	14	2	**10**	**93**
2	6	8	3	2	32	1	**119**	**100**
55	28	2	17	4	76	1	**492***	**105**
2	31	8	3	2	8	22	**532***	**111**
4	19	3	60	21	3	81	**498***	**112**
57	3	8	25	1	3	78	**477***	**113**

Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*) *=New types

Table.12. Genetic relatedness among *emm* types from community Impetigo and Normal throat

emm type	ST of Impetigo GAS	ST of Normal Throat GAS
11	471	471
28	474	474
63	338	338
66	249	249
82	320	320
85	109	109
100	119	119

Table.13. Genetic diversity among *emm* types from community Impetigo and Normal throat

***emm* type**	**ST of Impetigo GAS**	**ST of Normal Throat GAS**
1	15	530
44	178	477
55	100	528
92	100	485
105	492	491
111	532	496
112	498	497
113	477	500

Table.14.Allelic profile of GAS skin colonisers from community *(n=12)*

gki	***gtr***	***muri***	***muts***	***recp***	***xpt***	***yqil***	**ST**	***emm* type**
4	2	2	3	2	34	31	**535***	**18**
4	4	3	6	14	3	4	**178**	**44**
4	31	8	3	34	3	40	**534***	**49.4**
85	2	14	5	19	58	58	**338**	**63**
16	35	2	5	61	3	25	**481***	**75**
4	63	8	5	73	3	4	**320**	**82**
2	2	9	13	2	14	2	**10**	**93**
55	6	8	3	90	32	1	**488***	**100**
2	28	2	17	4	76	1	**491***	**105**
2	6	8	2	2	2	2	**533***	**108**
2	6	8	3	2	8	22	**496***	**111**
4	2	2	60	21	3	81	**497***	**112**

Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*)Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*) *=New types

Table.15. Distribution of allelic profiles of Hospital Pharyngitis GAS isolates*(n=20)*

gki	***gtr***	***muri***	***muts***	***recp***	***xpt***	***yqil***	**ST**	***emm*_type**
5	2	2	6	6	2	2	**36**	**1**
3	4	70	7	1	5	4	**471***	**11**
4	3	4	4	4	2	4	**28**	**12**
5	2	2	6	6	2	2	**36**	**12**
4	4	3	6	14	3	4	**178**	**44**
24	4	3	6	14	3	4	**351**	**44**
4	31	8	11	34	3	40	**228**	**49.4**
34	2	2	21	1	29	16	**100**	**55**
85	2	14	5	19	58	58	**338**	**63**
103	37	2	24	4	2	79	**479***	**73**
75	2	44	4	66	3	1	**357**	**75**
4	31	2	11	34	3	21	**482***	**77**
4	31	2	11	34	3	2	**347**	**77**
42	37	8	22	33	3	4	**109**	**85**
4	3	2	59	2	2	4	**485***	**92**
2	2	9	13	2	14	2	**10**	**93**
29	32	8	5	48	5	25	**501***	**118**
104	80	2	7	35	2	2	**483***	**st212**
4	36	12	18	2	3	4	**508***	**stKNB3.2**
29	31	6	5	19	77	4	**510***	**stNS554.1**

Glucose kinases (gki), Glutamine transporter protein (gtr), Glutamate racemase (murI), DNA mismatch repair protein (*mutS*)Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*) , *=New types

Table.16. Allelic profiles of Invasive blood Hospital GAS isolates*(n=32)*

source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* - types
blood	51	2	8	25	33	3	75	**470***	**8**
blood	4	28	2	3	2	34	31	**523***	**18**
blood	1	31	8	6	36	53	4	**475***	**43.3**
blood	4	3	3	6	14	3	77	**476***	**44**
blood	4	2	8	11	34	3	40	**478***	**49.4**
blood	90	6	8	7	4	3	59	**363**	**53**
blood	34	2	2	21	1	29	16	**100**	**55**
blood	85	2	14	5	19	58	58	**338**	**63**
blood	43	2	2	7	1	3	80	**480***	**74**
blood	43	32	2	7	1	3	1	**520***	**74**
blood	79	37	44	36	51	8	25	**521***	**81.2**
blood	42	2	8	22	33	3	4	**484***	**85**
blood	42	2	8	22	33	3	4	**484***	**85**
blood	4	12	36	11	4	23	1	**518***	**86**
blood	4	2	36	5	4	23	1	**503***	**86**
blood	2	6	9	13	2	14	2	**502***	**93**
blood	2	3	8	3	2	32	1	**487***	**100**
blood	2	12	8	3	2	32	1	**486***	**100**
blood	13	45	71	4	76	3	6	**489***	**102**
blood	25	37	8	7	2	3	1	**353**	**104**
blood	2	31	2	17	4	5	1	**490***	**105**
blood	25	37	8	6	30	25	4	**522***	**110**
blood	25	4	8	6	30	25	4	**493***	**110**
blood	25	3	8	6	30	25	4	**494***	**110**
blood	4	2	2	60	21	3	81	**497***	**112**
blood	29	31	2	5	48	5	25	**511***	**118**
blood	51	49	8	25	33	3	27	**544***	**124**
blood	51	2	2	25	2	5	31	**504***	**st1731**
blood	105	6	2	6	1	49	4	**506***	**st6735**
blood	105	3	2	6	1	49	4	**507***	**st6735**
blood	4	2	54	6	34	8	44	**519***	**st854.1**
blood	4	31	8	5	34	3	40	**517***	**stknb6**

*=New types, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table17. Allelic profiles of Invasive – Sterile body fluids of Hospital GAS isolates*(n=18*)

source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* types
CSF	48	31	41	9	2	2	36	**193**	**60**
CSF	13	31	8	5	19	3	1	**249**	**66**
CSF	43	2	2	7	1	3	80	**480***	**74**
CSF	4	59	2	5	44	33	1	**237**	**119**
CSF	51	49	8	25	33	3	27	**544***	**124**
fluid(necrotising fascitis)	2	28	2	17	4	5	1	**151**	**105**
fluid bile	25	2	8	6	30	25	39	**515***	**110**
Peritoneal fluid	14	2	8	6	55	40	6	**350**	**25**
Peritoneal fluid	4	36	8	7	51	3	76	**473***	**28**
Peritoneal fluid	85	2	14	5	19	58	58	**338**	**63**
Peritoneal fluid	4	3	54	6	34	8	44	**224**	**st854.1**
Peritoneal fluid	4	3	54	6	34	8	44	**224**	**st854.1**
Peritoneal fluid	4	36	2	18	2	3	4	**114**	**stKNB9**
Pleural fluid	47	3	2	5	10	64	12	**513***	**15**
Pleural fluid	40	2	9	6	41	34	12	**512***	**56**
synovial fluid	29	32	2	5	33	5	4	**516***	**58.8**
Pleural fluid	42	37	8	22	33	3	63	**514***	**85**
Pleural fluid	13	45	71	4	76	3	6	**489***	**102**

*=New types, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table.18.Comparison of STs from Community pharyngitis and Hospital pharyngitis

emm Type	STs of Community Pharyngitis	STs of Hospital Pharyngitis
1	469	36
11	-	471
12	472	28,36
18	526	-
28	474	-
44	-	178,351
49.4	534	228
55	100	100
63	338	338
73	-	479
75	-	357
77	482	347,482
82	320	-
85	525	109
92	524	485
93	10	10
105	151	-
108	527	-
110	495	-
112	497	-
113	499	-
118	-	501
st212	-	483
stNS554	510	510
stKNB8	509	-
stKNB3	-	508

Highlighted STs indicate same ST in the two groups

Table.19. Same ST type but different *emm* type

ST	*emm* Type
36	1,12
100	55,92
357	75,108
477	44,113

Table.20.Site of isolation of GAS isolates having same ST and different *emm* type

ST 36	***emm* 1**	***emm* 12**
	Hospital Pharyngitis	Impetigo
ST 357	***emm* 75**	***emm* 108**
	Hospital Pharyngitis	Normal Throat
ST 477	***emm* 44**	***emm* 113**
	Normal Throat	Impetigo
ST 100	***emm* 55**	***emm* 92**
	Community Pharyngitis	Impetigo
	Impetigo	
	Hospital Pharyngitis	
	Invasive	

Table.21.Distribution of diverse sequence types of same *emm* types based on their site of isolation

emm Type	Community Pharyngitis	Normal Throat	Impetigo	Normal Skin	Hospital Pharyngitis	blood	fluid
110	495	-	-	-	-	493,494,522	515
49.4	**534**	529	531	**534**	228	478	-
1	469	530	15	-	36	-	-
18	526	120	-	535	-	523	-
44	-	477	**178**	**178**	**178**,351	476	-
105	151	**491**	492	**491**	-	490	151
100	-	**119**	**119**	488	-	486,487	-
85	525	**109**	**109**	-	**109**	**484,484**	514
92	524	**485**	100	-	**485**	-	-
113	499	500	477	-	-	-	-
108	527	357	-	433	-	-	-
12	472	-	**36**	-	28,**36**	-	-
55	**100**	528	**100**	-	**100**	**100**	-
86	-	-	-	-	-	503,518	-
st6735	-	-	-	-	-	506,507	-
118	-	-	-	-	501	511	-
112	**497**	**497**	498	**497**	-	**497**	-
111	-	**496**	532	**496**	-	-	-
st212	-	505	-	-	483	-	-
77	**482**	**482**	-	-	347,**482**	-	-
75	-	-	-	481	357	-	-
28	**474**	**474**	**474**	-	-	-	473
st854.1	-	-	-	-	-	519	**224,224**
93	**10**	-	**10**	**10**	**10**	502	-
74	-	-	-	-	-	**480**,520	**480**

Highlighted STs indicate same ST among different site of isolation

Table .22. Allelic variations in *emm* type 1

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
IMP	2	6	8	5	2	3	2	**15**	**1**
HP	5	2	2	6	6	2	2	**36**	**1**
CP	26	2	2	5	31	3	1	**469**	**1**
NT	4	25	4	4	4	2	4	**530**	**1**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .23. Allelic variations in *emm* type 12

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
HP	4	3	4	4	4	2	4	**28**	**12**
HP	5	2	2	6	6	2	2	**36**	**12**
IMP	5	2	2	6	6	2	2	**36**	**12**
CP	4	2	2	58	6	2	2	**472**	**12**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, IMP = impetigo, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .24. Allelic variations in *emm* type 18

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
NT	43	2	2	7	1	3	1	**120**	**18**
Blood	4	28	2	3	2	34	31	**523**	**18**
CP	5	2	2	3	2	34	31	**526**	**18**
NS	4	2	2	3	2	34	31	**535**	**18**

CP = Community Pharyngitis, NS= Normal skin ,NT = Normal throat ,Glucose kinase (*gk* HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo *i*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .25. Allelic variations in *emm* type 44

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
HP	4	4	3	6	14	3	4	**178**	**44**
IMP	4	4	3	6	14	3	4	**178**	**44**
NS	4	4	3	6	14	3	4	**178**	**44**
HP	24	4	3	6	14	3	4	**351**	**44**
Blood	4	3	3	6	14	3	77	**476**	**44**
NT	57	3	8	25	1	3	78	**477**	**44**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo, NS= Normal skin Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .26. Allelic variations in *emm* type 49.4

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
HP	4	31	8	11	34	3	40	**228**	**49.4**
Blood	4	2	8	11	34	3	40	**478**	**49.4**
NT	4	6	8	3	34	3	40	**529**	**49.4**
IMP	4	35	8	3	34	3	40	**531**	**49.4**
CP	4	31	8	3	34	3	40	**534**	**49.4**
NS	4	31	8	3	34	3	40	**534**	**49.4**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo, NS= Normal skin Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .27 Allelic variations in *emm* type 85

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
HP	42	37	8	22	33	3	4	**109**	**85**
NT	42	37	8	22	33	3	4	**109**	**85**
IMP	42	37	8	22	33	3	4	**109**	**85**
Blood	42	2	8	22	33	3	4	**484**	**85**
Blood	42	2	8	22	33	3	4	**484**	**85**
PF	42	37	8	22	33	3	63	**514**	**85**
CP	4	37	8	22	33	3	4	**525**	**85**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo, NS= Normal skin, PF= Plueral fluid Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .28. Allelic variations in *emm* type 92

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
IMP	34	2	2	21	1	29	16	**100**	**92**
HP	4	3	2	59	2	2	4	**485**	**92**
NT	4	3	2	59	2	2	4	**485**	**92**
CP	42	3	2	5	2	2	4	**524**	**92**

CP = Community Pharyngitis, HP = Hospital Pharyngitis, NT = Normal throat, IMP = impetigo Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .29. Allelic variations in *emm* type 100

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
NT	2	6	8	3	2	32	1	**119**	**100**
IMP	2	6	8	3	2	32	1	**119**	**100**
Blood	2	12	8	3	2	32	1	**486**	**100**
Blood	2	3	8	3	2	32	1	**487**	**100**
NS	55	6	8	3	90	32	1	**488**	**100**

CP = Community Pharyngitis, NT = Normal throat, IMP = impetigo, NS= Normal skin, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .30. Allelic variations in *emm* type 105

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
NF	2	28	2	17	4	5	1	**151**	**105**
CP	2	28	2	17	4	5	1	**151**	**105**
Blood	2	31	2	17	4	5	1	**490**	**105**
NT	2	28	2	17	4	76	1	**491**	**105**
NS	2	28	2	17	4	76	1	**491**	**105**
IMP	55	28	2	17	4	76	1	**492**	**105**

CP = Community Pharyngitis, NT = Normal throat, IMP = impetigo, NS= Normal skin,NF=Necrotising fascitis Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .31. Allelic variations in *emm* type 108

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
NT	75	2	44	4	66	3	1	**357**	**108**
CP	2	6	3	22	2	2	2	**527**	**108**
NS	2	6	8	2	2	2	2	**533**	**108**

CP = Community Pharyngitis,NT = Normal throat,NS = Normal skin, Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*), Transketolase (*recP*), Acetyl-coA acetyl transferase (*Yqil*) ,Xanthine phosphoribosyl transferase (*xpt*).

Table .32. Allelic variations in *emm* type 110

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
Blood	25	4	8	6	30	25	4	**493**	**110**
Blood	25	3	8	6	30	25	4	**494**	**110**
CP	25	81	8	6	30	25	39	**495**	**110**
Bile	25	2	8	6	30	25	39	**515**	**110**
Blood	25	37	8	6	30	25	4	**522**	**110**

CP = Community Pharyngitis Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table .33. Allelic variations in *emm* type 113

Source	*gki*	*gtr*	*muri*	*muts*	*recp*	*xpt*	*yqil*	ST	*emm* type
IMP	57	3	8	25	1	3	78	**477**	**113**
CP	57	3	8	25	1	3	76	**499**	**113**
CP	57	3	8	25	1	3	76	**499**	**113**
NT	57	3	8	25	19	3	78	**500**	**113**

CP = Community Pharyngitis Glucose kinase (*gki*), Glutamine transporter protein (*gtr*), Glutamate racemase (*murI*), DNA mismatch repair protein (*mutS*) Transketolase (*recP*), Xanthine phosphoribosyl transferase (*xpt*), Acetyl-coA acetyl transferase (*Yqil*)

Table.34.Score Table of maximum nucleotide identity of *emm* 1

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST15	3134	2	ST36	3134	99
1	ST15	3134	3	ST469	3134	99
1	ST15	3134	4	ST530	3134	99
2	ST36	3134	3	ST469	3134	99
2	ST36	3134	4	ST530	3134	99
3	ST469	3134	4	ST530	3134	99

Table.35.Score Table of maximum nucleotide identity of *emm* 12

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST528	3134	2	ST36	3134	99
1	ST528	3134	3	ST472	3134	99
2	ST36	3134	3	ST472	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.36.Score Table of maximum nucleotide identity of *emm* 18

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST120	3134	2	ST523	3134	99
1	ST120	3134	3	ST526	3134	99
1	ST120	3134	4	ST535	3134	99
2	ST523	3134	3	ST526	3134	99
2	ST523	3134	4	ST535	3134	99
3	ST526	3134	4	ST535	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.37.Score Table of maximum nucleotide identity of *emm* 44

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST178	3134	2	ST351	3134	99
1	ST178	3134	3	ST476	3134	99
1	ST178	3134	4	ST477	3134	99
2	ST351	3134	3	ST476	3134	99
2	ST351	3134	4	ST477	3134	99
3	ST476	3134	4	ST477	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.38.Score Table of maximum nucleotide identity of *emm* 49.4

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST478	3134	2	ST228	3134	99
1	ST478	3134	3	ST534	3134	99
1	ST478	3134	4	ST531	3134	99
1	ST478	3134	5	ST529	3134	99
2	ST228	3134	3	ST534	3134	99
2	ST228	3134	4	ST531	3134	99
2	ST228	3134	5	ST529	3134	99
3	ST534	3134	4	ST531	3134	99
3	ST534	3134	5	ST529	3134	99
4	ST531	3134	5	ST529	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.39.Score Table of maximum nucleotide identity of *emm* 85

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST109	3134	2	ST484	3134	99
1	ST109	3134	3	ST514	3134	99
1	ST109	3134	4	ST525	3134	99
2	ST484	3134	3	ST514	3134	99
2	ST484	3134	4	ST525	3134	99
3	ST514	3134	4	ST525	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.40.Score Table of maximum nucleotide identity of *emm* 92

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST100	3134	2	ST485	3134	99
1	ST100	3134	3	ST524	3134	99
2	ST485	3134	3	ST524	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.41.Score Table of maximum nucleotide identity of *emm* 100

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST119	3134	2	ST486	3134	99
1	ST119	3134	3	ST487	3134	99
1	ST119	3134	4	ST488	3134	99
2	ST486	3134	3	ST487	3134	99
2	ST486	3134	4	ST488	3134	99
3	ST487	3134	4	ST488	3134	99

Table.42.Score Table of maximum nucleotide identity of *emm* 105

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST490	3134	2	ST491	3134	99
1	ST490	3134	3	ST492	3134	99
1	ST490	3134	4	ST151	3134	99
2	ST491	3134	3	ST492	3134	99
2	ST491	3134	4	ST151	3134	99
3	ST492	3134	4	ST151	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.43.Score Table of maximum nucleotide identity of *emm* 108

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST357	3134	2	ST527	3134	98
1	ST357	3134	3	ST533	3134	98
2	ST527	3134	3	ST533	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.44.Score Table of maximum nucleotide identity of *emm* 110

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST522	3134	2	ST515	3134	99
1	ST522	3134	3	ST493	3134	99
1	ST522	3134	4	ST494	3134	99
2	ST515	3134	3	ST493	3134	99
2	ST515	3134	4	ST494	3134	99
3	ST493	3134	4	ST494	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

Table.45.Score Table of maximum nucleotide identity of *emm* 113

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ST477	3134	2	ST499	3134	99
1	ST477	3134	3	ST500	3134	99
2	ST499	3134	3	ST500	3134	99

Seq A = Sequence A, Seq B = Sequence B, Len (nt) = Nucleotide length

10. BIBLIOGRAPHY

1. Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev 2000;13:470-511

2. wannamaker. treatment of the preceeding streptococcal infection. J amer med assoc 1950;143:151-153

3. Shet A, Kaplan EL, Johnson DR and Cleary PP. Immune response to group A streptococcal C5a peptidase in children: implications for vaccine development. J Infect Dis 2003;188:809-17

4. Veasy LG. resurgence of acute RF in the intermountain area of US. N Engl J Med 1987;316:421-427

5. acute RF in western pennsylvania and thr tri state area. Pedatrics 1987;18:371-374

6. Veasey LG. persistence of acute RF in intermountain area of US. J Pediatr 1994;124:9-16

7. Stevens DL. Streptococcal toxic-shock syndrome: spectrum of disease, pathogenesis, and new concepts in treatment. Emerg Infect Dis 1995;1:69-78

8. lannigan R. an electron microscope study of mucopolysaccharide in RH leision. 1968;96:305-309

9. Kaplan D. presence of bound immunoglobulins and complement in the myocardium in ARF. N Engl J Med 1964;271:637-645

10. Kaplan D. immunological studies of heart tissue occurence of boud gamma globulin in aricular appendages from RH. J Exp Med 1961;113:1-16

11. Seigel. heart reactive antibody associated with RF. clin Exp Immunol 1970;7:147-159

12. Carapetis JR, Steer AC, Mulholland EK and Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis 2005;5:685-94

13. Carapetis JR. Rheumatic heart disease in developing countries. N Engl J Med 2007;357:439-41

14. Kaplan EL. Epidemiological approaches to understanding the pathogenesis of rheumatic fever. Int J Epidemiol 1985;14:499-501

15. Cunningham MW, Antone SM, Gulizia JM, McManus BA and Gauntt CJ. Alpha-helical coiled-coil molecules: a role in autoimmunity against the heart. Clin Immunol Immunopathol 1993;68:118-23

16. Padmavathi Sa. Prevalence odf Rheumatic heart disease in delhi school children. Indian J of medical research 1979;69:821

17. Grace Koshi VBaGC. Rheumatic fever and rheumatic heart disease in rural South Indian children. Bulletin of the World Health Organisation 1981;59(4):599-603

18. M.Myers GKaR. Streptococcal disease in children in Southern India. The Indian Journal of Pathology and Bacteriology 1971;14:17-23

19. Padmavati S. The cardiac patient in underdeveloped countries. Am Heart J 1959;58:418-24

20. Sagar V, Bakshi DK, Nandi S, Ganguly NK, Kumar R and Chakraborti A. Molecular heterogeneity among north Indian isolates of Group A Streptococcus. Lett Appl Microbiol 2004;39:84-8

21. Brahmadathan KN, Gladstone P. Microbiological diagnosis of streptococcal pharyngitis: lacunae and their implications. Indian J Med Microbiol 2006;24:92-6

22. Bisno AL, Gerber MA, Gwaltney JM, Jr., Kaplan EL and Schwartz RH. Diagnosis and management of group A streptococcal pharyngitis: a practice guideline. Infectious Diseases Society of America. Clin Infect Dis 1997;25:574-83

23. Medina E, Chhatwal GS. The potential for vaccine development against rheumatic fever. Indian Heart J 2002;54:93-8

24. Macris MH, Hartman N, Murray B, et al. Studies of the continuing susceptibility of group A streptococcal strains to penicillin during eight decades. Pediatr Infect Dis J 1998;17:377-81

25. Markowitz M. Streptococcal disease in developing countries. Pediatr Infect Dis J 1991;10:S11-4

26. McMillan DJ, Sriprakash KS and Chhatwal GS. Genetic variation in group A streptococci. Int J Med Microbiol 2007;297:525-32

27. Bessen DE, Fiorentino TR and Hollingshead SK. Molecular markers for throat and skin isolates of group A streptococci. Adv Exp Med Biol 1997;418:537-43

28. Musser JM, Kapur V, Kanjilal S, et al. Geographic and temporal distribution and molecular characterization of two highly pathogenic clones of Streptococcus pyogenes

expressing allelic variants of pyrogenic exotoxin A (Scarlet fever toxin). J Infect Dis 1993;167:337-46

29. Beall B, Facklam R, Hoenes T and Schwartz B. Survey of emm gene sequences and T-antigen types from systemic Streptococcus pyogenes infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. J Clin Microbiol 1997;35:1231-5

30. Chiou CS, Liao TL, Wang TH, Chang HL, Liao JC and Li CC. Epidemiology and molecular characterization of Streptococcus pyogenes recovered from scarlet fever patients in central Taiwan from 1996 to 1999. J Clin Microbiol 2004;42:3998-4006

31. Facklam R, Beall B, Efstratiou A, et al. emm typing and validation of provisional M types for group A streptococci. Emerg Infect Dis 1999;5:247-53

32. Beall B, Facklam R and Thompson T. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J Clin Microbiol 1996;34:953-8

33. Beall B, Facklam RR, Elliott JA, et al. Streptococcal emm types associated with T-agglutination types and the use of conserved emm gene restriction fragment patterns for subtyping group A streptococci. J Med Microbiol 1998;47:893-8

34. Yakubu DE, Abadi FJ and Pennington TH. Molecular typing methods for Neisseria meningitidis. J Med Microbiol 1999;48:1055-64

35. Wisplinghoff H, Hippler C, Bartual SG, et al. Molecular epidemiology of clinical Acinetobacter baumannii and Acinetobacter genomic species 13TU isolates using a multilocus sequencing typing scheme. Clin Microbiol Infect 2008;14:708-15

36. Robinson DA, Sutcliffe JA, Tewodros W, Manoharan A and Bessen DE. Evolution and global dissemination of macrolide-resistant group A streptococci. Antimicrob Agents Chemother 2006;50:2903-11

37. Aanensen DM, Spratt BG. The multilocus sequence typing network: mlst.net. Nucleic Acids Res 2005;33:W728-33

38. Enright MC, Spratt BG, Kalia A, Cross JH and Bessen DE. Multilocus sequence typing of Streptococcus pyogenes and the relationships between emm type and clone. Infect Immun 2001;69:2416-27

39. Horiike T, Miyata D, Hamada K, et al. Phylogenetic construction of 17 bacterial phyla by new method and carefully selected orthologs. Gene 2009;429:59-64

40. Chiou CS, Liao JC, Liao TL, et al. Molecular epidemiology and emergence of worldwide epidemic clones of Neisseria meningitidis in Taiwan. BMC Infect Dis 2006;6:25

41. Spratt BG, Hanage WP, Li B, Aanensen DM and Feil EJ. Displaying the relatedness among isolates of bacterial species -- the eBURST approach. FEMS Microbiol Lett 2004;241:129-34

42. Coenye T, LiPuma JJ. Population structure analysis of Burkholderia cepacia genomovar III: varying degrees of genetic recombination characterize major clonal complexes. Microbiology 2003;149:77-88

43. Scheet P, Stephens M. Linkage disequilibrium-based quality control for large-scale genetic studies. PLoS Genet 2008;4:e1000147

44. Smith JM SN, O'Rourke M, Spratt BG. How clonal are bacteria? Proc Natl Acad Sci 1993;90:4384-8

45. Kalia A, Spratt BG, Enright MC and Bessen DE. Influence of recombination and niche separation on the population genetic structure of the pathogen Streptococcus pyogenes. Infect Immun 2002;70:1971-83

46. Maxted WR. Group A streptococci: pathogenesis and immunity. Soc Appl Bacteriol Symp Ser 1978;7:107-25

47. Mann AH. IV. A Case of Emphysematous Gangrene of the Hand Due to the Streptococcus Pyogenes and the Bacillus AErogenes Capsulatus: Recovery without Amputation. Ann Surg 1894;19:187-96

48. Holmes MC, Williams RE. The distribution of carriers of Streptococcus pyogenes among 2,413 healthy children. J Hyg (Lond) 1954;52:165-79

49. Evans AC. Studies on Hemolytic Streptococci: II. Streptococcus pyogenes. J Bacteriol 1936;31:611-24

50. Rountree PM. Streptococcus pyogenes infections in a hospital. Lancet 1955;269:172-3

51. Mc CM, Lancefield RC. Variation in the group-specific carbohydrate of group A streptococci. I. Immunochemical studies on the carbohydrates of variant strains. J Exp Med 1955;102:11-28

52. Shaunak S, Wendon J, Monteil M and Gordon AM. Septic scarlet fever due to Streptococcus pyogenes cellulitis. Q J Med 1988;69:921-5

53. Conterno G, Agostoni A, Capra A, Colongo PG and Ramello A. [Antigens and antibodies of Streptococcus pyogenes in the laboratory and in clinical medicine. Clinico-experimental research on ASO, ASK, ASH and anti-M antibodies]. Arch Sci Med (Torino) 1967;124:672-715

54. Watson RF, Hirst GK and Lancefield RC. Bacteriological studies of cardiac tissues obtained at autopsy from eleven patients dying with rheumatic fever. Arthritis Rheum 1961;4:74-85

55. Rotta J. [Long-term examination of the prevalence of the group and type (group A) of Streptococcus pyogenes in Czechoslovakia. Study in the summer of 1960-1961.]. Cesk Epidemiol Mikrobiol Imunol 1963;12:65-73

56. Ludwig W SE, Klipper-Balz R. The phylogenetic Position of *Streptococcus* and *Enterococcus*. J Gen Microbiol 1985;131:543

57. Ferretti JJ, McShan WM, Ajdic D, et al. Complete genome sequence of an M1 strain of Streptococcus pyogenes. Proc Natl Acad Sci U S A 2001;98:4658-63

58. Stephen B.Beres JMM. Contribution of Exogenous Genetic Elements to the Group A Streptococcus Metagenome. PLoS ONE 2007;2

59. Lancefield DE. serological differentiation of human and other groups of beta hemolytic streptococci. J Exp Med 1933;57:571-595

60. Sherman JM, Albus WR. Some Characters which Differentiate the Lactic-acid Streptococcus from Streptococci of the Pyogenes Type Occurring in Milk. J Bacteriol 1918;3:153-74

61. Schleifer KH K-B. Molecular and chemotaxonomic approaches to the classification of streptococci. Syst Appl Microbiol 1987;10:1-9

62. Kawamura Y, Hou XC, Sultana F. determination of 16srRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationship among members of the genus *Streptococcus*. Syst Bacteriol 1995;45:406-408

63. Pike. An enrichment broth for isolating hemolytic streptococci from throat swabs. Proceedings of the society for Experimental Biology and Medicine 1944;57

64. Saito M, Ohga S, Endoh M, et al. H(2)O(2)-nonproducing Streptococcus pyogenes strains: survival in stationary phase and virulence in chronic granulomatous disease. Microbiology 2001;147:2469-77

65. Koshi G, Mammen KC. Hemolytic streptococci and immune response in children with rheumatic fever. Indian J Med Res 1969;57:1347-60

66. White AG, Steele RH and Pierce WA, Jr. The effect of pH on the fermentation of glucose and galactose by Streptococcus pyogenes. J Bacteriol 1955;70:82-6

67. Plummer DT, James AM, Gooder JH and Maxted WR. Some physical investigations of the behaviour of bacterial surfaces. V. The variation of the surface structure of Streptococcus pyogenes during growth. Biochim Biophys Acta 1962;60:595-603

68. Chapman S. Unusual Group A Streptococci: Colonial appearance and hemolysis. In : Wannamaker LW, Maxtsen JM,eds.*streptococci andstreptococcal disease: recognition, understanding and management*. Newyork , Aadamic press 1972:211-233

69. Kocka FE. Nutritionally Variant *Streptococcus pyogenes* from a periorbital abscess. Journal of Clinical Microbiology 1987;25:736-737

70. Chapman S, Johnson DR, Dudding BA. Group A beta hemolytic Streptococci growing as satellite colonies Bacteriological proceedings 1969:72

71. RM P. An enrichment broth for isolating hemolytic streptococci from throat swabs. Proceedings of the society for Experimental Biology and Medicine 1944;57:186-187

72. PD E. A new cultual medium for medical bacteriology. American Journal of clinical pathology 1966;45:502-504

73. DN P. Colistin- Oxolinic acid - blood agar : a new selective medim for streptococci Journal of Clinical Microbiology 1984;19:4-7

74. Gunn BA. Selective and enhanced recovery of group A and B streptococci from throat cultures with sheep blood agar containing sulfamethoxazole and trimethoprim. Journal of Clinical Microbiology 1977;5:650-655

75. Tolliver PR RM, Todd JK. Detection of Group A streptococcus: Comparison of solid and liquid culture media with and with out selective antibiotics . Pediatric Infectious disease journal 1987;6:515-519

76. Koshi G, Brahmadathan KN, Pandian R and Joseph A. Modified filter paper strip technique for transportation & rapid isolation of streptococci from throat swabs. Indian J Med Res 1988;88:336-8

77. Redys JJ HE, Borman EK. Improved dry swab transportation for streptococcal specimens. Public health reports 1968;83:143-149

78. Koshi G, Brahmadathan KN, Thangavelu CP and Pandian R. Evaluation of different methods for the transport of swabs for streptococci. Indian J Med Res 1979;69:26-31

79. Malathy B Tm. Evaluation of crysal violet impregnated filter paper strips for transport and selective isolation of beta hemolytic streptococci. Journal of Medical Microbiology 2007

80. Koshi G, Rajeshwari K and Philipose L. Preservation of streptococci and other bacteria by sand desiccation and filter paper techniques. Indian J Med Res 1977;65:500-2

81. Bronze MS, McKinsey DS, Beachey EH and Dale JB. Pathogenesis of group A streptococci in mice and efficacy of locally administered streptococcal vaccines. Trans Assoc Am Physicians 1988;101:88-92

82. Wessels M R, and M S Bronze. Critcal role of the Group A streptococcal capsule in pharyngeal colonisation and infection in mice. Proc.Natl.Acd.Sci 1994;91:12238-12242

83. Maxted WR. The M and T antigens of Streptococcus pyogenes type 2. J Pathol Bacteriol 1953;65:345-54

84. Baum H, Crowley N. Starch-likepolysaccharide of Streptococcus pyogenes. Nature 1960;187:413

85. Cole RM, Hahn JJ. Cell wall replication in Streptococcus pyogenes. Science 1962;135:722-4

86. W.CUNNINGHAM M. Pathogenesis of Group A Streptococcal Infections. Clinical Microbiology Reviews 2000;13:470-511

87. Bisno AL, Brito MO and Collins CM. Molecular basis of group A streptococcal virulence. Lancet Infect Dis 2003;3:191-200

88. Courtney HS, Hasty DL and Dale JB. Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. Ann Med 2002;34:77-87

89. Beachey EH, Simpson WA. The adherence of group A streptococci to oropharyngeal cells: the lipoteichoic acid adhesin and fibronectin receptor. Infection 1982;10:107-11

90. Fischetti VA. Streptococcal M protein extracted by nonionic detergent. III. Correlation between immunological cross-reactions and structural similarities with implications for antiphagocytosis. J Exp Med 1978;147:1771-8

91. Hill MJ, James AM and Maxted WR. Some physical investigations of the behaviour of bacterial surfaces. VIII. Studies on the capsular material of Streptococcus pyogenes. Biochim Biophys Acta 1963;66:264-74

92. Dinkla K, Sastalla I, Godehardt AW, et al. Upregulation of capsule enables Streptococcus pyogenes to evade immune recognition by antigen-specific antibodies directed to the G-related alpha2-macroglobulin-binding protein GRAB located on the bacterial surface. Microbes Infect 2007;9:922-31

93. Schrager HM, Alberti S, Cywes C, Dougherty GJ and Wessels MR. Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A Streptococcus to CD44 on human keratinocytes. J Clin Invest 1998;101:1708-16

94. Dougherty BA, and I. Van de Rijn. Molecular characterisation of a locus required for Hyaluronic acid capsule production in Group A streptococci. J Exp Med 1992;175:1291-1299

95. Aashbaugh CD AS, Wessels M Molecular analysis of the capsule gene region of Group A streptococci : the *has* AB genes are sufficient for capsule expression. J Bacteriol 1988;180

96. whitnack .E BA, Beachey EH. Hyaluronate capsule prevents attachment of Group A streptococci to peritoneal macrophages. Infect.Immun 1981;31:985-991

97. Fillit H, M. McCarty.and M . Blake. lnduction of antibodies to hyaluronic acid by immunization of rabbits with encapsulated streptococci. J Exp Med 1986;164:762-766

98. Benchetrit LC, Wannamaker LW and Gray ED. Immunological properties of hyaluronidases associated with temperate bacteriophages of group A streptococci. J Exp Med 1979;149:73-83

99. Johnson DR, Stevens DL and Kaplan EL. Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever, or uncomplicated pharyngitis. J Infect Dis 1992;166:374-82

100. Greenblatt J, Boackle RJ and Schwab JH. Activation of the alternate complement pathway by peptidoglycan from streptococcal cell wall. Infect Immun 1978;19:296-303

101. Heymann H, Manniello JM and Barkulis SS. Structure of streptococcal cell walls. V. Phosphate esters in the walls of group A Streptococcus pyogenes. Biochem Biophys Res Commun 1967;26:486-91

102. Bisno AL. Molecular aspects of bacterial colonization. Infect Control Hosp Epidemiol 1995;16:648-57

103. Courtney HS, Ofek I, Penfound T, et al. Relationship between expression of the family of M proteins and lipoteichoic acid to hydrophobicity and biofilm formation in Streptococcus pyogenes. PLoS ONE 2009;4:e4166

104. Koga T, Kakimoto K, Hirofuji T, et al. Acute joint inflammation in mice after systemic injection of the cell wall, its peptidoglycan, and chemically defined peptidoglycan subunits from various bacteria. Infect Immun 1985;50:27-34

105. Jarva H, Jokiranta TS, Wurzner R and Meri S. Complement resistance mechanisms of streptococci. Mol Immunol 2003;40:95-107

106. Beachey EH, Dale JB, Simpson WA, et al. Erythrocyte binding properties of streptococcal lipoteichoic acids. Infect Immun 1979;23:618-25

107. Rotta J. Endotoxin-like properties of the peptidoglycan. Z Immunitatsforsch Exp Klin Immunol 1975;149:230-44

108. Alouf J, Muller-Alouf H. [Cellular constituents and extracellular proteins involved in the pathogenic capacity of Streptococcus group A]. Ann Pharm Fr 1996;54:49-59

109. Talay SR, Zock A, Rohde M, et al. Co-operative binding of human fibronectin to Sfbl protein triggers streptococcal invasion into respiratory epithelial cells. Cell Microbiol 2000;2:521-35

110. Alkan M, Ofek I and Beachey EH. Adherence pharyngeal and skin strains of group A streptococci to human skin and oral epithelial cells. Infect Immun 1977;18:555-7

111. Talay SR, Valentin-Weigand P, Jerlstrom PG, Timmis KN and Chhatwal GS. Fibronectin-binding protein of Streptococcus pyogenes: sequence of the binding domain involved in adherence of streptococci to epithelial cells. Infect Immun 1992;60:3837-44

112. Rocha CL, Fischetti VA. Identification and characterization of a new protein from Streptococcus pyogenes having homology with fibronectin and fibrinogen binding proteins. Adv Exp Med Biol 1997;418:737-9

113. Hyland KA, Wang B and Cleary PP. Protein F1 and Streptococcus pyogenes resistance to phagocytosis. Infect Immun 2007;75:3188-91

114. McArthur J, Medina E, Mueller A, et al. Intranasal vaccination with streptococcal fibronectin binding protein Sfb1 fails to prevent growth and dissemination of Streptococcus pyogenes in a murine skin infection model. Infect Immun 2004;72:7342-5

115. McArthur J, Schulze K, Chin J, et al. Immune responses of a liposome/ISCOM vaccine adjuvant against streptococcal fibronectin binding protein 1 (Sfb1) in mice. Indian J Med Res 2004;119 Suppl:115-20

116. Guzman CA, Talay SR, Molinari G, Medina E and Chhatwal GS. Protective immune response against Streptococcus pyogenes in mice after intranasal vaccination with the fibronectin-binding protein SfbI. J Infect Dis 1999;179:901-6

117. Flores AE, Johnson DR, Kaplan EL and Wannamaker LW. Factors influencing antibody responses to streptococcal M proteins in humans. J Infect Dis 1983;147:1-15

118. Beachey EH, Seyer JM and Kang AH. Repeating covalent structure of streptococcal M protein. Proc Natl Acad Sci U S A 1978;75:3163-7

119. Fischetti VA, Jones KF, Manjula BN and Scott JR. Streptococcal M6 protein expressed in Escherichia coli. Localization, purification, and comparison with streptococcal-derived M protein. J Exp Med 1984;159:1083-95

120. Beachey EH, Seyer JM, Dale JB and Hasty DL. Repeating covalent structure and protective immunogenicity of native and synthetic polypeptide fragments of type 24 streptococcal M protein. Mapping of protective and nonprotective epitopes with monoclonal antibodies. J Biol Chem 1983;258:13250-7

121. Beachey EH, Stollerman GH, Chiang EY, Chiang TM, Seyer JM and Kang AH. Purification and properties of M protein extracted from group A streptococci with pepsin: covalent structure of the amino terminal region of type 24 M antigen. J Exp Med 1977;145:1469-83

122. Fischetti VA. Streptococcal M protein. Sci Am 1991;264:58-65

123. Katsukawa C, Harada K and Makino M. [M-typing of Streptococcus pyogenes by latex agglutination using monoclonal antibody. Streptococcal Diseases Study Group]. Kansenshogaku Zasshi 1994;68:706-11

124. Nilay dey DJM, Penny J.Yarwood,R.M.Joshi,Rajesh Kumar,Michael F.Good,Kadaba S. Sriprakash and Harpreet Vohra. High Diversity of Group A streptococcal *emm* types in a Indian Community: The need to Tailor Multivalent Vaccines. Clinical Infectious disease 2005;40:46-51

125. Sakota V, Fry AM, Lietman TM, Facklam RR, Li Z and Beall B. Genetically diverse group A streptococci from children in far-western Nepal share high genetic relatedness with isolates from other countries. J Clin Microbiol 2006;44:2160-6

126. Noble RC, Penny BB. Characterization of the antibody response to type 12 M protein of group A streptococcus. J Lab Clin Med 1975;86:713-21

127. Dale JB, Ofek I and Beachey EH. Heterogeneity of type-specific and cross-reactive antigenic determinants within a single M protein of group A streptococci. J Exp Med 1980;151:1026-38

128. Wittner MK, Fox EN. Homologous and heterologous protection of mice with group A streptococcal M protein vaccines. Infect Immun 1977;15:104-8

129. Bessen D, Jones KF and Fischetti VA. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. J Exp Med 1989;169:269-83

130. Courtney HS, Hasty DL and Dale JB. Serum opacity factor (SOF) of Streptococcus pyogenes evokes antibodies that opsonize homologous and heterologous SOF-positive serotypes of group A streptococci. Infect Immun 2003;71:5097-103

131. Bessen DE, Veasy LG, Hill HR, Augustine NH and Fischetti VA. Serologic evidence for a class I group A streptococcal infection among rheumatic fever patients. J Infect Dis 1995;172:1608-11

132. D'Alessandri R, Plotkin G, Kluge RM, et al. Protective studies with group A streptococcal M protein vaccine. III. Challenge of volunteers after systemic or intranasal immunization with Type 3 or Type 12 group A Streptococcus. J Infect Dis 1978;138:712-8

133. Castagnetta E, Milazzo M and Merucci P. [Cardiotoxicity of a Cellular Component of Beta-Hemolytic Streptococcus, Group A.]. Rend Ist Sup Sanit 1963;26:1088-101

134. Hollingshead SK, Bessen DE. Evolution of the emm gene family: virulence gene clusters in group A streptococci. Dev Biol Stand 1995;85:163-8

135. Hollingshead SK, Readdy TL, Yung DL and Bessen DE. Structural heterogeneity of the emm gene cluster in group A streptococci. Mol Microbiol 1993;8:707-17

136. Hollingshead SK, Arnold J, Readdy TL and Bessen DE. Molecular evolution of a multigene family in group A streptococci. Mol Biol Evol 1994;11:208-19

137. Bessen DE, Sotir CM, Readdy TL and Hollingshead SK. Genetic correlates of throat and skin isolates of group A streptococci. J Infect Dis 1996;173:896-900

138. Fagan PK, Reinscheid D, Gottschalk B and Chhatwal GS. Identification and characterization of a novel secreted immunoglobulin binding protein from group A streptococcus. Infect Immun 2001;69:4851-7

139. Jeppson H, Frithz E and Heden LO. Duplication of a DNA sequence homologous to genes for immunoglobulin receptors and M proteins in Streptococcus pyogenes. FEMS Microbiol Lett 1992;71:139-45

140. Heath DG, Cleary PP. Cloning and expression of the gene for an immunoglobulin G Fc receptor protein from a group A streptococcus. Infect Immun 1987;55:1233-8

141. Podbielski A, Flosdorff A and Weber-Heynemann J. The group A streptococcal virR49 gene controls expression of four structural vir regulon genes. Infect Immun 1995;63:9-20

142. Frithz E, Heden LO and Lindahl G. Extensive sequence homology between IgA receptor and M proteins in Streptococcus pyogenes. Mol Microbiol 1989;3:1111-9

143. Haanes EJ, Cleary PP. Identification of a divergent M protein gene and an M protein-related gene family in Streptococcus pyogenes serotype 49. J Bacteriol 1989;171:6397-408

144. Heath DG, Cleary PP. Fc-receptor and M-protein genes of group A streptococci are products of gene duplication. Proc Natl Acad Sci U S A 1989;86:4741-5

145. Top FH, Jr., Wannamaker LW. The serum opacity reaction of Streptococcus pyogenes: frequency of production of streptococcal lipoproteinase by strains of different serological types and the relationship of M protein production. J Hyg (Lond) 1968;66:49-58

146. Beall B, Gherardi G, Lovgren M, Facklam RR, Forwick BA and Tyrrell GJ. emm and sof gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. Microbiology 2000;146 (Pt 5):1195-209

147. Kreikemeyer B, Talay SR and Chhatwal GS. Characterization of a novel fibronectin-binding surface protein in group A streptococci. Mol Microbiol 1995;17:137-45

148. Courtney HS, Hasty DL, Li Y, Chiang HC, Thacker JL and Dale JB. Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 Streptococcus pyogenes. Mol Microbiol 1999;32:89-98

149. Katerov V, Lindgren PE, Totolian AA and Schalen C. Streptococcal opacity factor: a family of bifunctional proteins with lipoproteinase and fibronectin-binding activities. Curr Microbiol 2000;40:149-56

150. Rakonjac JV, Robbins JC and Fischetti VA. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. Infect Immun 1995;63:622-31

151. Courtney HS, Dale JB and Hasty DL. Mapping the fibrinogen-binding domain of serum opacity factor of group a streptococci. Curr Microbiol 2002;44:236-40

152. Kreikemeyer B, Martin DR and Chhatwal GS. SfbII protein, a fibronectin binding surface protein of group A streptococci, is a serum opacity factor with high serotype-specific apolipoproteinase activity. FEMS Microbiol Lett 1999;178:305-11

153. Courtney HS, Hasty DL and Dale JB. Anti-phagocytic mechanisms of Streptococcus pyogenes: binding of fibrinogen to M-related protein. Mol Microbiol 2006;59:936-47

154. Oehmcke S, Podbielski A and Kreikemeyer B. Function of the fibronectin-binding serum opacity factor of Streptococcus pyogenes in adherence to epithelial cells. Infect Immun 2004;72:4302-8

155. Jeng A, Sakota V, Li Z, Datta V, Beall B and Nizet V. Molecular genetic analysis of a group A Streptococcus operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. J Bacteriol 2003;185:1208-17

156. Cleary PP, Prahbu U, Dale JB, Wexler DE and Handley J. Streptococcal C5a peptidase is a highly specific endopeptidase. Infect Immun 1992;60:5219-23

157. Chen CC, Cleary PP. Complete nucleotide sequence of the streptococcal C5a peptidase gene of Streptococcus pyogenes. J Biol Chem 1990;265:3161-7

158. Podbielski A, Kaufhold A and Lutticken R. [The vir-regulon of Streptococcus pyogenes: coordinate expression of important virulence factors]. Immun Infekt 1992;20:161-8

159. Ji Y, McLandsborough L, Kondagunta A and Cleary PP. C5a peptidase alters clearance and trafficking of group A streptococci by infected mice. Infect Immun 1996;64:503-10

160. Ji Y, Schnitzler N, DeMaster E and Cleary P. Impact of M49, Mrp, Enn, and C5a peptidase proteins on colonization of the mouse oral mucosa by Streptococcus pyogenes. Infect Immun 1998;66:5399-405

161. Ji Y, Carlson B, Kondagunta A and Cleary PP. Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A Streptococcus. Infect Immun 1997;65:2080-7

162. O'Connor SP, Darip D, Fraley K, Nelson CM, Kaplan EL and Cleary PP. The human antibody response to streptococcal C5a peptidase. J Infect Dis 1991;163:109-16

163. Wang H, Lottenberg R and Boyle MD. A role for fibrinogen in the streptokinase-dependent acquisition of plasmin(ogen) by group A streptococci. J Infect Dis 1995;171:85-92

164. Rezcallah MS, Boyle MD and Sledjeski DD. Mouse skin passage of Streptococcus pyogenes results in increased streptokinase expression and activity. Microbiology 2004;150:365-71

165. Wang JR, Stinson MW. M protein mediates streptococcal adhesion to HEp-2 cells. Infect Immun 1994;62:442-8

166. Tillet WaRLG. The fibrinolytic activity of hemolytic streptococci. J Exp Med 1933;58:485-502

167. Huang TT, Malke H and Ferretti JJ. Heterogeneity of the streptokinase gene in group A streptococci. Infect Immun 1989;57:502-6

168. Walter F, Siegel M and Malke H. Nucleotide sequence of the streptokinase gene from a Streptococcus pyogenes type 1 strain. Nucleic Acids Res 1989;17:1261

169. Ohkuni H, Todome Y, Suzuki H, et al. Immunochemical studies and complete amino acid sequence of the streptokinase from Streptococcus pyogenes (group A) M type 12 strain A374. Infect Immun 1992;60:278-83

170. Ben Nasr A, Wistedt A, Ringdahl U and Sjobring U. Streptokinase activates plasminogen bound to human group C and G streptococci through M-like proteins. Eur J Biochem 1994;222:267-76

171. Gerlach D, Kohler W. [Studies of the heterogeneity of streptokinases; IV. Evidence for isostreptokinases in Streptococcus pyogenes type 1 (author's transl)]. Zentralbl Bakteriol A 1981;248:446-54

172. Tewodros W, Nordstrand A, Kronvall G, Holm SE and Norgren M. Streptokinase gene polymorphism in group A streptococci isolated from Ethiopian children with various disease manifestations. Microb Pathog 1993;15:303-11

173. Churchward G, Bates C, Gusa AA, Stringer V and Scott JR. Regulation of streptokinase expression by CovR/S in Streptococcus pyogenes: CovR acts through a single high-affinity binding site. Microbiology 2009;155:566-75

174. Oberley TD, Duncan JL. Characteristics of streptolysin O action. Infect Immun 1971;4:683-7

175. Limbago B, Penumalli V, Weinrick B and Scott JR. Role of streptolysin O in a mouse model of invasive group A streptococcal disease. Infect Immun 2000;68:6384-90

176. Ayoub EM, Nelson B, Shulman ST, et al. Group A streptococcal antibodies in subjects with or without rheumatic fever in areas with high or low incidences of rheumatic fever. Clin Diagn Lab Immunol 2003;10:886-90

177. Dale JB, Chiang EY, Hasty DL and Courtney HS. Antibodies against a synthetic peptide of SagA neutralize the cytolytic activity of streptolysin S from group A streptococci. Infect Immun 2002;70:2166-70

178. Lee PK, Schlievert PM. Quantification and toxicity of group A streptococcal pyrogenic exotoxins in an animal model of toxic shock syndrome-like illness. J Clin Microbiol 1989;27:1890-2

179. Johnson LP, L'Italien JJ and Schlievert PM. Streptococcal pyrogenic exotoxin type A (scarlet fever toxin) is related to Staphylococcus aureus enterotoxin B. Mol Gen Genet 1986;203:354-6

180. Goshorn SC, Schlievert PM. Bacteriophage association of streptococcal pyrogenic exotoxin type C. J Bacteriol 1989;171:3068-73

181. Weeks CR, Ferretti JJ. The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. Infect Immun 1984;46:531-6

182. Weeks CR, Ferretti JJ. Nucleotide sequence of the type A streptococcal exotoxin (erythrogenic toxin) gene from Streptococcus pyogenes bacteriophage T12. Infect Immun 1986;52:144-50

183. Yu CE, Ferretti JJ. Molecular characterization of new group A streptococcal bacteriophages containing the gene for streptococcal erythrogenic toxin A (speA). Mol Gen Genet 1991;231:161-8

184. Goshorn SC, Schlievert PM. Nucleotide sequence of streptococcal pyrogenic exotoxin type C. Infect Immun 1988;56:2518-20

185. Hauser AR, Schlievert PM. Nucleotide sequence of the streptococcal pyrogenic exotoxin type B gene and relationship between the toxin and the streptococcal proteinase precursor. J Bacteriol 1990;172:4536-42

186. Yu CE, Ferretti JJ. Frequency of the erythrogenic toxin B and C genes (speB and speC) among clinical isolates of group A streptococci. Infect Immun 1991;59:211-5

187. Kapur V, Majesky MW, Li LL, Black RA and Musser JM. Cleavage of interleukin 1 beta (IL-1 beta) precursor to produce active IL-1 beta by a conserved extracellular cysteine protease from Streptococcus pyogenes. Proc Natl Acad Sci U S A 1993;90:7676-80

188. Kapur V, Topouzis S, Majesky MW, et al. A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin. Microb Pathog 1993;15:327-46

189. Kamezawa Y, Nakahara T, Nakano S, Abe Y, Nozaki-Renard J and Isono T. Streptococcal mitogenic exotoxin Z, a novel acidic superantigenic toxin produced by a T1 strain of Streptococcus pyogenes. Infect Immun 1997;65:3828-33

190. Iwasaki M, Igarashi H, Hinuma Y and Yutsudo T. Cloning, characterization and overexpression of a Streptococcus pyogenes gene encoding a new type of mitogenic factor. FEBS Lett 1993;331:187-92

191. Sriskandan S, Faulkner L and Hopkins P. Streptococcus pyogenes: Insight into the function of the streptococcal superantigens. Int J Biochem Cell Biol 2007;39:12-9

192. Nelson K, Schlievert PM, Selander RK and Musser JM. Characterization and clonal distribution of four alleles of the speA gene encoding pyrogenic exotoxin A (scarlet fever toxin) in Streptococcus pyogenes. J Exp Med 1991;174:1271-4

193. Kline JB, Collins CM. Analysis of the superantigenic activity of mutant and allelic forms of streptococcal pyrogenic exotoxin A. Infect Immun 1996;64:861-9

194. Gerlach D, Knoll H, Kohler W, Ozegowski JH and Hribalova V. Isolation and characterization of erythrogenic toxins. V. Communication: identity of erythrogenic toxin type B and streptococcal proteinase precursor. Zentralbl Bakteriol Mikrobiol Hyg [A] 1983;255:221-33

195. Musser JM, Stockbauer K, Kapur V and Rudgers GW. Substitution of cysteine 192 in a highly conserved Streptococcus pyogenes extracellular cysteine protease (interleukin 1beta convertase) alters proteolytic activity and ablates zymogen processing. Infect Immun 1996;64:1913-7

196. Holm SE, Norrby A, Bergholm AM and Norgren M. Aspects of pathogenesis of serious group A streptococcal infections in Sweden, 1988-1989. J Infect Dis 1992;166:31-7

197. Chaussee MS, Liu J, Stevens DL and Ferretti JJ. Effects of environmental factors on streptococcal erythrogenic toxin A (SPE A) production by Streptococcus pyogenes. Adv Exp Med Biol 1997;418:551-4

198. Svensson MD, Sjobring U, Luo F and Bessen DE. Roles of the plasminogen activator streptokinase and the plasminogen-associated M protein in an experimental model for streptococcal impetigo. Microbiology 2002;148:3933-45

199. Aziz RK, Ismail SA, Park HW and Kotb M. Post-proteomic identification of a novel phage-encoded streptodornase, Sda1, in invasive M1T1 Streptococcus pyogenes. Mol Microbiol 2004;54:184-97

200. Wang CH, Lin CY, Luo YH, et al. Effects of oligopeptide permease in group a streptococcal infection. Infect Immun 2005;73:2881-90

201. Abuhammour W, Hasan RA and Unuvar E. Group A beta-hemolytic streptococcal bacteremia. Indian J Pediatr 2004;71:915-9

202. Koshi G, Jadhav M and Myers RM. Streptococcal pharyngitis in children. Indian J Med Res 1970;58:161-7

203. Bisno AL. Group A streptococcal infections and acute rheumatic fever. N Engl J Med 1991;325:783-93

204. Bisno AL, Stevens DL. Streptococcal infections of skin and soft tissues. N Engl J Med 1996;334:240-5

205. Invasive streptococci. Lancet 1989;2:1255

206. Georgiev P, Vurleva T. [Puerperal streptococcal sepsis]. Akush Ginekol (Sofiia) 1993;32:3-5

207. Stevens DL, Tanner MH, Winship J, et al. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. N Engl J Med 1989;321:1-7

208. Bahna M, Canalis RF. Necrotizing fasciitis. (Streptococcal gangrene) of the face. Report of a case and review of the literature. Arch Otolaryngol 1980;106:648-51

209. Stevens DL. Invasive group A streptococcus infections. Clin Infect Dis 1992;14:2-11

210. Defining the group A streptococcal toxic shock syndrome. Rationale and consensus definition. The Working Group on Severe Streptococcal Infections. JAMA 1993;269:390-1

211. McDonald M, Currie BJ and Carapetis JR. Acute rheumatic fever: a chink in the chain that links the heart to the throat? Lancet Infect Dis 2004;4:240-5

212. Shulman ST. Rheumatic heart disease in developing countries. N Engl J Med 2007;357:2089; author reply 2089

213. Yoshizawa N. Acute Glomerulonephritis. Intern Med 2000;39

214. Coleman DJ, McGhie D and Tebbutt GM. Further studies on the reliability of the bacitracin inhibition test for the presumptive identification of Lancefield group A streptococci. J Clin Pathol 1977;30:421-6

215. Johnson DR, E. Kaplan, J. Sramek, R. Bicova, J. Havlicek, H. Havlickova,J. Motlova, and P. Kriz. Laboratory methods for the diagnosis of group A streptococcal infections. 1996 World Health Organization,Geneva, Switzerland.)

216. Sridharan G, Thangavelu CP, Sivadasan K, Vanitha M, Rajeswari P and Koshi G. Grouping of beta-haemolytic streptococci from primary plates by micronitrous acid-coagglutination method. Indian J Med Res 1984;79:479-81

217. Lieu TA, Fleisher GR and Schwartz JS. Clinical evaluation of a latex agglutination test for streptococcal pharyngitis: performance and impact on treatment rates. Pediatr Infect Dis J 1988;7:847-54

218. Diagnosis and treatment of streptococcal pharyngitis. JAMA 1985;254:907-9

219. Koshi G, Sridharan G and Thangavelu CP. Antibody to group A streptococcal carbohydrate in rheumatic fever and rheumatic heart disease. Indian J Med Res 1983;77:172-7

220. Koshi G. Anti-streptolysin O test by microtechnique. Indian J Med Res 1971;59:1554-5

221. Brahmadathan KN, Anitha P and Gladstone P. Increasing erythromycin resistance among group A streptococci causing tonsillitis in a tertiary care hospital in southern India. Clin Microbiol Infect 2005;11:335-7

222. Kaplan EL, Cornaglia G. Persistent macrolide resistance among group A streptococci: the lack of accomplishment after 4 decades. Clin Infect Dis 2005;41:609-11

223. Johnson DR, Kaplan EL, VanGheem A, Facklam RR and Beall B. Characterization of group A streptococci (Streptococcus pyogenes): correlation of M-protein and emm-gene

type with T-protein agglutination pattern and serum opacity factor. J Med Microbiol 2006;55:157-64

224. Creti R, Cardona F, Pataracchia M, et al. Characterisation of group A streptococcal (GAS) isolates from children with tic disorders. Indian J Med Res 2004;119 Suppl:174-8

225. Dicuonzo G, Gherardi G, Lorino G, et al. Group A streptococcal genotypes from pediatric throat isolates in Rome, Italy. J Clin Microbiol 2001;39:1687-90

226. von Hunolstein C, Suligoi B, Pataracchia M, et al. Clinical and microbiological characteristics of severe group A streptococcal infections in Italy. Adv Exp Med Biol 1997;418:79-81

227. Creti R, Gherardi G, Imperi M, et al. Association of group A streptococcal emm types with virulence traits and macrolide-resistance genes is independent of the source of isolation. J Med Microbiol 2005;54:913-7

228. Bessen DE, Carapetis JR, Beall B, et al. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. J Infect Dis 2000;182:1109-16

229. O'Brien KL, Beall B, Barrett NL, et al. Epidemiology of invasive group a streptococcus disease in the United States, 1995-1999. Clin Infect Dis 2002;35:268-76

230. Davies HD, McGeer A, Schwartz B, et al. Invasive group A streptococcal infections in Ontario, Canada. Ontario Group A Streptococcal Study Group. N Engl J Med 1996;335:547-54

231. Vlaminckx BJ, Mascini EM, Schellekens J, et al. Site-specific manifestations of invasive group a streptococcal disease: type distribution and corresponding patterns of virulence determinants. J Clin Microbiol 2003;41:4941-9

232. Bart J.M. Vlamnickx EMM, Joop Schellekens,Leo M.Schouls, Armand Paauw, Ad C Fluit, Rodger Nanak, Jan Verhoef, and Franz Josef Schmitz. Site-Specific Manifestations of Invasive Group A Streptococcal Disease: Type Distribution and Corresponding Patterns of Virulence Determinants. Journal of Clinical Microbiology 2003;41:4941-4949

233. Kim SJ, Kim EC, Cha SH and Kaplan EL. Comparison of M-serotypes of Streptococcus pyogenes isolated from healthy elementary school children in two rural areas. J Korean Med Sci 1996;11:133-6

234. Mejia LM, Stockbauer KE, Pan X, Cravioto A and Musser JM. Characterization of group A Streptococcus strains recovered from Mexican children with pharyngitis by automated DNA sequencing of virulence-related genes: unexpectedly large variation in the gene (sic) encoding a complement-inhibiting protein. J Clin Microbiol 1997;35:3220-4

235. Nguyen L, Levy D, Ferroni A, Gehanno P and Berche P. Molecular epidemiology of Streptococcus pyogenes in an area where acute pharyngotonsillitis is endemic. J Clin Microbiol 1997;35:2111-4

236. Gardiner DL, Sriprakash KS. Molecular epidemiology of impetiginous group A streptococcal infections in aboriginal communities of northern Australia. J Clin Microbiol 1996;34:1448-52

237. Martin MA, Hebden JN, Bustamante CI and Rubin LJ. Group A streptococcal bacteremias associated with intravascular catheters. Infect Control Hosp Epidemiol 1990;11:542-4

238. Schwartz B, Facklam RR and Breiman RF. Changing epidemiology of group A streptococcal infection in the USA. Lancet 1990;336:1167-71

239. Ho PL, Johnson DR, Yue AW, et al. Epidemiologic analysis of invasive and noninvasive group a streptococcal isolates in Hong Kong. J Clin Microbiol 2003;41:937-42

240. Cleary PP, Kaplan EL, Handley JP, et al. Clonal basis for resurgence of serious Streptococcus pyogenes disease in the 1980s. Lancet 1992;339:518-21

241. Johnson DR, Wotton JT, Shet A and Kaplan EL. A comparison of group A streptococci from invasive and uncomplicated infections: are virulent clones responsible for serious streptococcal infections? J Infect Dis 2002;185:1586-95

242. Chaussee MS, Liu J, Stevens DL and Ferretti JJ. Genetic and phenotypic diversity among isolates of Streptococcus pyogenes from invasive infections. J Infect Dis 1996;173:901-8

243. Mylvaganam H, Bjorvatn B, Hofstad T and Osland A. Molecular characterization and allelic distribution of the phage-mediated hyaluronidase genes hylP and hylP2 among group A streptococci from western Norway. Microb Pathog 2000;29:145-53

244. Rogers S, Commons R, Danchin MH, et al. Strain prevalence, rather than innate virulence potential, is the major factor responsible for an increase in serious group A streptococcus infections. J Infect Dis 2007;195:1625-33

245. Bessen DE, Izzo MW, Fiorentino TR, Caringal RM, Hollingshead SK and Beall B. Genetic linkage of exotoxin alleles and emm gene markers for tissue tropism in group A streptococci. J Infect Dis 1999;179:627-36

246. Kalia A, Bessen DE. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. J Bacteriol 2004;186:110-21

247. Tewodros W, Kronvall G. M protein gene (emm type) analysis of group A beta-hemolytic streptococci from Ethiopia reveals unique patterns. J Clin Microbiol 2005;43:4369-76

248. Gardiner D, Hartas J, Currie B, Mathews JD, Kemp DJ and Sriprakash KS. Vir typing: a long-PCR typing method for group A streptococci. PCR Methods Appl 1995;4:288-93

249. Caparon MG, Scott JR. Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. Proc Natl Acad Sci U S A 1987;84:8677-81

250. Gardiner D, Hartas J, Hibble M, Goodfellow A, Currie B and Sriprakash KS. Molecular epidemiology of group A streptococcal infection in the Northern Territory of Australia. Adv Exp Med Biol 1997;418:317-21

251. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998;95:3140-5

252. McGregor KF, Bilek N, Bennett A, et al. Group A streptococci from a remote community have novel multilocus genotypes but share emm types and housekeeping alleles with isolates from worldwide sources. J Infect Dis 2004;189:717-23

253. McGregor KF, Spratt BG. Identity and prevalence of multilocus sequence typing-defined clones of group A streptococci within a hospital setting. J Clin Microbiol 2005;43:1963-7

254. Reinert RR, Lutticken R, Sutcliffe JA, et al. Clonal relatedness of erythromycin-resistant Streptococcus pyogenes isolates in Germany. Antimicrob Agents Chemother 2004;48:1369-73

255. Roger Meisal EAH, Inerborge S. Aaberge and Dominique A.Caugant. Sequence Type and *emm* Type Diversity in *Streptococcus pyogenes*. Isolate Causing Disease in Norway between 1988 and 2003. Journal of Clinical Microbiology 2008;46:2102-2105

256. Benouda A, Sibile S, Ziane Y, Elouennass M, Dahani K and Hassani A. [Place of Streptococcus pyogenes in the throat infections and overview of its susceptibility to antibiotics]. Pathol Biol (Paris) 2009;57:76-80

257. Fischetti VA, Bessen DE, Schneewind O and Hruby DE. Protection against streptococcal pharyngeal colonization with vaccines composed of M protein conserved regions. Adv Exp Med Biol 1991;303:159-67

258. Fischetti VA, Hodges WM and Hruby DE. Protection against streptococcal pharyngeal colonization with a vaccinia: M protein recombinant. Science 1989;244:1487-90

259. Bessen D, Fischetti VA. Synthetic peptide vaccine against mucosal colonization by group A streptococci. I. Protection against a heterologous M serotype with shared C repeat region epitopes. J Immunol 1990;145:1251-6

260. Cunningham MW. Pathogenesis of group A streptococcal infections and their sequelae. Adv Exp Med Biol 2008;609:29-42

261. McMillan DJ, Chhatwal GS. Prospects for a group A streptococcal vaccine. Curr Opin Mol Ther 2005;7:11-6

262. Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B and Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. Infect Immun 2002;70:2171-7

263. Feil EJ, Li BC, Aanensen DM, Hanage WP and Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 2004;186:1518-30

264. Kumar S, Nei M, Dudley J and Tamura K. MEGA: biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 2008;9:299-306

265. Tamura K, Dudley J, Nei M and Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596-9

266. Smith JM, Smith NH, O'Rourke M and Spratt BG. How clonal are bacteria? Proc Natl Acad Sci U S A 1993;90:4384-8

267. Thompson JD, Higgins DG and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-80

268. Vivek Sagar RK, Nirmal K Ganguly and Anuradha chakraborti. Comparative analysis of emm type pattern of Group A Streptococcus throat and skin isolates from India and their

association with closely related SIC, a streptococcal virulence factor. BMC Microbiol 2008;8

269. Thangam Menon CL, Malathy B,Varja Sakota, Delois Jackson and Beall. emm type diversity of beta hemolytic streptococci recovered in India. JMM 2007

270. McGregor KF, Spratt BG, Kalia A, et al. Multilocus sequence typing of Streptococcus pyogenes representing most known emm types and distinctions among subpopulation genetic structures. J Bacteriol 2004;186:4285-94

271. Turner KM, Hanage WP, Fraser C, Connor TR and Spratt BG. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol 2007;7:30

272. Kumar S, Dudley J. Bioinformatics software for biologists in the genomics era. Bioinformatics 2007;23:1713-7